

# Estimation of a semiparametric contaminated regression model

Pierre Vandekerkhove

*LAMA - CNRS UMR 8050,*

*Université Paris-Est*

*Pierre.Vandekerkhove@univ-mlv.fr*

## Abstract

We consider in this paper a contaminated regression model where the distribution of the contaminating component is known when the Euclidean parameters of the regression model, the noise distribution, the contamination ratio and the distribution of the design data are unknown. Our model is said to be semiparametric in the sense that the probability density function (pdf) of the noise involved in the regression model is not supposed to belong to a parametric density family. When the pdf's of the noise and the contaminating phenomenon are supposed to be symmetric about zero, we propose an estimator of the various (Euclidean and fonctionnal) parameters of the model, and prove under mild conditions its convergence. We prove in particular that, under technical conditions all satisfied in the Gaussian case, the Euclidean part of the model is estimated at the rate  $o_{a.s.}(n^{-1/4+\gamma})$ ,  $\gamma > 0$ . We recall that, as it is pointed out in Bordes and Vandekerkhove [5], this result cannot be ignored to go further in the asymptotic theory for this class of models. Finally the implementation and numerical performances of our method are discussed on several toy examples.

**Keywords.** M-estimator, mixture, regression model, empirical process, semiparametric identifiability, uniform convergence rate.

# 1 Introduction

Let  $(U_i)_{i \geq 1}$  be a sequence of independent and identically distributed (iid) random variables according to a *Bernoulli* distribution with parameter  $p \in (0, 1)$ . We consider an iid sample  $(Z_1, \dots, Z_n)$  where for all  $i = 1, \dots, n$ ,  $Z_i = (X_i, Y_i)$  is a bivariate random variable defined, relative to  $U_i$ , as follows

$$\begin{cases} Y_i = a_0 + b_0 X_i + \varepsilon_i^{[0]}, & \text{if } U_i = 0, \\ Y_i = a_1 + b_1 X_i + \varepsilon_i^{[1]}, & \text{if } U_i = 1, \end{cases} \quad (1)$$

where the design sequence  $(X_i)_{i \geq 1}$ , respectively the errors  $(\varepsilon_i^{[j]})_{i \geq 1}$ ,  $j = 0, 1$ , is a sequence of iid random variables with cumulative distribution function (cdf)  $H$ , resp.  $F_j$ , and probability density function (pdf),  $h$ , resp.  $f_j$ ,  $j = 0, 1$ . We suppose in addition that the design sequence is independent from the errors. This model, called the *2-mixture of regression model*, belongs to the wide class of mixture of regression models which has been studied in [29]; see also [26] in a LOS (length of stay) medical problem, [6] for prediction, or [27] in a nonparametric modelling context. Recently Martin-Magniette *et al.* [21] introduced this model in microarray analysis for the study of the two color ChIP-chip experiment. Briefly, the Chromatin immunoprecipitation (ChIP) is a well established procedure to investigate proteins associated with DNA. ChIP on chip involves analysis of DNA recovered from ChIP experiments by hybridization to microarray. In a two color ChIP-chip experiment, two samples are compared: DNA fragments crosslinked to a protein of interest (IP) and genomic DNA (input). The goal is then to identify actual binding targets of the IP, *i.e.* probes whose IP signal is significantly larger than the input signal. In the model proposed by Martin-Magniette *et al.* [21] the components of the random vector  $Z_i = (X_i, Y_i)$ , see model (1), corresponds respectively to the log-input and log-IP intensities of probe  $i$  when the (unknown) status of the probe is characterized through a label  $U_i$  which is 1 if the probe is *enriched* and 0 if it is *standard* (not enriched). Note also that the assumption made by these authors on the error sequences  $(\varepsilon_i^{[j]})_{i \geq 1}$ ,  $j = 0, 1$ , is that  $\varepsilon_i^{[j]} = \varepsilon_i$  for all  $(j, i) \in \{0, 1\} \times \mathbb{N}^*$  where  $\varepsilon_i$  is a Gaussian random variable with mean 0 and variance  $\sigma^2$  (homoscedasticity with respect to the probe status  $U_i$ ).

In this work, we propose to weaken this last assumption while completely specifying the regression model under the probe standard condition (the parameter  $\theta^{[0]} := (a_0, b_0) \in \mathbb{R}^2$  and  $f_0$  are supposed to be entirely known). Note that this kind of assumption arises naturally in microarray analysis, see model (5) and references [1], [12], or [3] p. 744 formula (22), where analytic expression of  $f_0$ , characterizing probe expressivity levels under a certain standard condition, is assumed to be available (generally derived from training data and probabilistic computations). In particular we will suppose that, in model (1), the distribution of the  $\varepsilon_i^{[1]}$  is seen as a nuisance parameter (it is no longer supposed to belong to a parametric distribution family), turning model (1) into a purely semiparametric model. Note that when  $\theta^{[0]}$  is known the observations  $Y_i$ , for  $i = 1, \dots, n$ , can be centered according to  $Y_i := Y_i - (a_0 + b_0 X_i)$  which implies a simplification of model (1), since we then have

$$\begin{cases} Y_i = \varepsilon_i^{[0]}, & \text{if } U_i = 0, \\ Y_i = \alpha + \beta X_i + \varepsilon_i^{[1]}, & \text{if } U_i = 1, \end{cases} \quad (2)$$

where  $\alpha := a_1 - a_0$  and  $\beta := b_1 - b_0$ . We suppose in model (2), which is from now on our model of interest, that the  $Z_i = (X_i, Y_i)$ 's distribution admits a pdf with respect to the Lebesgue measure on  $\mathbb{R}^2$  defined by:

$$\begin{aligned} g(x, y) &= h(x)g_{Y|X=x}(y) \\ &= h(x)[pf(y - (\alpha + \beta x)) + (1 - p)f_0(y)], \quad (x, y) \in \mathbb{R}^2, \end{aligned} \quad (3)$$

where  $f$  denotes the unknown pdf of the  $\varepsilon_i^{[1]}$ ,  $f_0$  the known pdf of the  $\varepsilon_i^{[0]}$ ,  $h$  the unknown pdf of the  $X_i$ ,  $f$  and  $f_0$  being supposed to belong to the class of even densities. We will finally denote by  $\vartheta := (p, \alpha, \beta) \in (0, 1) \times \mathbb{R}^2$  the unknown Euclidean parameter of model (3). Model (2) corresponds exactly to a contaminated version of the semiparametric additive regression model studied in [9], [10] and more recently in [28]. On the other hand model (3) extends for the first time to the bivariate case, the class of semiparametric mixture models introduced by Hall and Zhou [13] for  $\mathbb{R}^s$ -valued observations with  $s \geq 3$ , and studied later in the univariate case, through two specific models:

$$g(y) = pf(y - \mu_1) + (1 - p)f(y - \mu_2), \quad y \in \mathbb{R}, \quad (4)$$

where  $(p, \mu_1, \mu_2) \in (0, 1/2) \times \mathbb{R}^2$  and  $f$ , supposed to be even, are unknown, see [2], [17], [20], and

$$g(y) = pf(y) + (1 - p)f_0(y - \mu_2), \quad y \in \mathbb{R}, \quad (5)$$

where  $(p, \mu) \in (0, 1) \times \mathbb{R}$  and  $f$  are unknown,  $f_0$  is known, and the pdfs  $f$  and  $f_0$  are supposed to be even, see [3], [5].

The paper is organized as follows. In Section 2 we present an M-estimating method, inspired by [2], [3] and [5], that allows us to estimate the Euclidean and the functional parameters of model (2); in Section 3 we address the semi-parametric identifiability problem associated to expression (3) and establish rates of convergence of our estimators; in Section 4 we discuss the performance of our method on simulated examples and focus our attention on the optimization problems encountered during its implementation. When technical results are relegated to the appendix, which corresponds to Section 5.

## 2 Estimating method

In the spirit of [2], [3] and [5], we will suppose that  $f$  and  $f_0$  are both pdfs symmetric about zero (recall that only  $f_0$  is assumed known). To avoid trivial situations or trivial non-identifiability problems (see Remark in Section 3.1), we will impose  $p \neq 1$  and  $\theta := (\alpha, \beta) \in \Phi \subset \mathbb{R} \times \mathbb{R}^*$ , which implies that the Euclidean parameter  $\vartheta$  will be assumed to belong to a parametric compact and convex space

$$\Theta := [\delta, 1 - \delta] \times \Phi \subset (0, 1) \times \{\mathbb{R} \times \mathbb{R}^*\}, \quad (6)$$

where  $\delta \in (0, 1)$ .

For simplicity, we will endow the spaces  $\mathbb{R}^s$ ,  $s \geq 1$ , with the  $\|\cdot\|_s$  norm (for clarity the dimension  $s$  is recalled in index) defined for all  $v = (v_1, \dots, v_s)$  by  $\|v\|_s = \sum_{j=1}^s |v_j|$  where  $|\cdot|$  denotes the absolute value.

We now introduce the following non-commutative notation:

$$\theta \odot x := \alpha + \beta x, \quad (\theta, x) \in \Phi \times \mathbb{R}.$$

Following the ideas developed by the authors mentioned above, it is possible to use the symmetry assumption made on  $f$  to identify the true value of the Euclidean parameter. The idea consists in noticing that for  $\theta$  fixed in  $\Phi$ , the sample  $(Y_1^\theta, \dots, Y_n^\theta)$  obtained by considering the so-called  $\theta$ -transformation

$$Y_i^\theta := Y_i - \theta \odot X_i, \quad i = 1, \dots, n, \quad (7)$$

is distributed according to

$$\Psi_\theta(y) = p_* \int_{\mathbb{R}} f(y + (\theta - \theta_*) \odot x) h(x) dx + (1 - p_*) \int_{\mathbb{R}} f_0(y + \theta \odot x) h(x) dx, \quad (8)$$

where  $\vartheta_* = (p_*, \alpha_*, \beta_*) \in \mathring{\Theta}$  denotes the true value of the parameter. Let us observe now that when  $\theta = \theta_*$

$$\Psi_{\theta_*}(y) = p_* f(y) + (1 - p_*) \int_{\mathbb{R}} f_0(y + \theta_* \odot x) h(x) dx. \quad (9)$$

*Remark.* When  $\theta$  is well fitted ( $\theta = \theta_*$ ) the model associated to the  $Y^\theta$  is very close to the simple contamination model (5) studied in [3] or [5] where the location  $\mu$  is known but the proportion  $p$  is unknown.

Isolating  $f$  in (9) and replacing  $\vartheta_* = (p_*, \theta_*)$  by  $\vartheta = (p, \theta)$  one can define a new parametric class of functions  $\mathcal{F}_\Theta := \{f_\vartheta : \vartheta \in \Theta\}$ :

$$f_\vartheta(y) = \frac{1}{p} \Psi_\theta(y) - \frac{1-p}{p} \int_{\mathbb{R}} f_0(y + \theta \odot x) h(x) dx, \quad (y, \vartheta) \in \mathbb{R} \times \Theta \quad (10)$$

that satisfies under  $\vartheta = \vartheta_*$ ,

$$f(y) = f_{\vartheta_*}(y) = f_{\vartheta_*}(-y) = f(-y), \quad y \in \mathbb{R}. \quad (11)$$

The intuition consists now in claiming that, if we make  $\vartheta$  vary over  $\Theta$  and that we are able to check that  $f_\vartheta$  is symmetric about 0 for a certain value of  $\vartheta$  then we have reached the true value of the Euclidean parameter. Note that in the right hand side of (10), the second integral term is in general unknown but can be estimated pointwise by a standard Monte Carlo approach, see expression (16), or a nonparametric Monte Carlo approach, see expression (24). The idea

to check this situation, and then to estimate  $\vartheta = (p, \theta)$ , is to consider a contrast function based on the comparison between the cdf version of  $f_\vartheta(y)$

$$H_1(y; \vartheta) := H_1(y; p, F_\theta, J_\theta) := \frac{1}{p}F_\theta(y) - \frac{1-p}{p}J_\theta(y), \quad (y, \theta) \in \mathbb{R} \times \Theta,$$

and the cdf version of  $f_\vartheta(-y)$

$$H_2(y; \vartheta) := H_2(y; p, F_\theta, J_\theta) := 1 - \frac{1}{p}F_\theta(-y) + \frac{1-p}{p}J_\theta(-y), \quad (y, \theta) \in \mathbb{R} \times \Theta,$$

where for all  $\theta \in \Phi$ ,

$$J_\theta(y) := \int_{-\infty}^y I_\theta(z) dz, \quad y \in \mathbb{R}, \quad \text{with } I_\theta(z) := \int_{\mathbb{R}} f_0(z + \theta \odot x) h(x) dx, \quad z \in \mathbb{R},$$

and

$$F_\theta(y) := \int_{-\infty}^y f_\theta(z) dz, \quad y \in \mathbb{R}.$$

Notice that for all  $\theta$  fixed in  $\Phi$ ,  $J_\theta(\cdot)$  and  $F_\theta(\cdot)$  are the cdfs associated respectively to the  $\theta$ -transformed known component population (the  $Y_i$  such that  $U_i = 0$  in (2)) and the  $\theta$ -transformed whole data. Let us define the following function

$$H(y; \vartheta) := H_1(y; \vartheta) - H_2(y; \vartheta), \quad (y, \vartheta) \in \mathbb{R} \times \Theta. \quad (12)$$

Notice that under  $\vartheta_*$ , using the symmetry of  $f$ ,

$$H(y; \vartheta_*) = 0, \quad y \in \mathbb{R}.$$

To avoid numerical integration in the approximation of an empirical contrast function based on the comparison of  $H_1$  and  $H_2$  over  $\mathbb{R}$ , we proceed as follows. Let  $Q$  be an *instrumental* weight probability distribution with pdf  $q$  with respect to Lebesgue measure. We suppose that  $q$  is strictly positive over  $\mathbb{R}$  and easy to simulate. Then we consider

$$d(\vartheta) := \int_{\mathbb{R}} H^2(y, \vartheta) dQ(y), \quad (13)$$

where obviously  $d(\vartheta) \geq 0$  for all  $\vartheta \in \Theta$  and  $d(\vartheta_*) = 0$ . Let  $(V_1, \dots, V_n)$  be an iid sample from  $Q$ . An empirical version  $d_n(\cdot)$  of  $d(\cdot)$  can be obtained by considering

$$d_n(\vartheta) := \frac{1}{n} \sum_{i=1}^n H^2(V_i; p, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta}), \quad \vartheta \in \Theta, \quad (14)$$

where

$$\hat{J}_{n,\theta}(y) := \int_{-\infty}^y \hat{I}_{n,\theta}(z) dz, \quad (y, \theta) \in \mathbb{R} \times \Phi, \quad (15)$$

with

$$\hat{I}_{n,\theta}(z) := \frac{1}{n} \sum_{i=1}^n f_0(z + \theta \odot X_i), \quad (z, \theta) \in \mathbb{R} \times \Phi, \quad (16)$$

which leads actually to the simple expression for  $\hat{J}_{n,\theta}(y)$

$$\hat{J}_{n,\theta}(y) := \frac{1}{n} \sum_{i=1}^n F_0(y + \theta \odot X_i), \quad (y, \theta) \in \mathbb{R} \times \Phi, \quad (17)$$

and where  $\tilde{F}_{n,\theta}$  denotes a smooth version of the empirical cdf

$$\hat{F}_{n,\theta}(y) := \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{Y_i^\theta \leq y}, \quad (y, \theta) \in \mathbb{R} \times \Phi,$$

defined by

$$\tilde{F}_{n,\theta}(y) := \int_{-\infty}^y \hat{\Psi}_{n,\theta}(t) dt, \quad (y, \theta) \in \mathbb{R} \times \Phi, \quad (18)$$

where

$$\hat{\Psi}_{n,\theta}(t) := \frac{1}{nb_n} \sum_{i=1}^n K\left(\frac{t - Y_i^\theta}{b_n}\right), \quad (t, \theta) \in \mathbb{R} \times \Phi. \quad (19)$$

In (19), we assume the standard condition insuring, for each  $\theta \in \Phi$ , the  $L_1$  convergence of  $\hat{\Psi}_{n,\theta}$  towards  $\Psi_\theta$  defined in (8) (see Devroye [11]), namely

$$b_n \rightarrow 0, \quad nb_n \rightarrow +\infty, \quad (20)$$

and  $K$  is a symmetric density function. Finally we propose to estimate  $\vartheta_*$  by considering the M-estimator

$$\hat{\vartheta}_n := (\hat{p}_n, \hat{\theta}_n) = \arg \min_{\vartheta \in \Theta} d_n(\vartheta). \quad (21)$$

Once  $\vartheta_*$  is estimated by  $\hat{\vartheta}_n$  a natural way to estimate  $F$  and  $f$  consistently is then to consider the plug-in empirical versions of  $H_1(\cdot; \vartheta)$  and (10), respectively defined for all  $y \in \mathbb{R}$  by

$$\hat{F}_n(y) := H_1(y; \hat{p}_n, \tilde{F}_{n,\hat{\theta}_n}, \tilde{J}_{n,\hat{\theta}_n}), \quad (22)$$

$$\hat{f}_n(y) := \frac{1}{\hat{p}_n} \hat{\Psi}_{n,\hat{\theta}_n}(y) + \frac{1 - \hat{p}_n}{\hat{p}_n} \tilde{I}_{n,\hat{\theta}_n}(y), \quad (23)$$

where, for all  $\theta \in \Theta$ ,  $\tilde{I}_{n,\theta}$  and  $\tilde{J}_{n,\theta}$  are respectively nonparametric estimators of  $I_\theta$  and  $J_\theta$  based on an iid simulated sample  $(\tilde{\varepsilon}_1^{[0]}, \dots, \tilde{\varepsilon}_n^{[0]})$  from  $f_0$  obtained by considering

$$\tilde{I}_{n,\theta}(t) := \frac{1}{nb_n} \sum_{i=1}^n K \left( \frac{t - (\theta \odot X_i + \tilde{\varepsilon}_i^{[0]})}{b_n} \right), \quad (t, \theta) \in \mathbb{R} \times \Phi. \quad (24)$$

$$\tilde{J}_{n,\theta}(y) := \int_{-\infty}^y \tilde{I}_{n,\theta}(t) dt, \quad (y, \theta) \in \mathbb{R} \times \Phi, \quad (25)$$

For convenience, the kernel used to compute (24) will be Gaussian, *i.e.*  $K(t) = \mathcal{N}_{0,1}(t)$  where  $\mathcal{N}_{m,\sigma^2}(t) := (2\pi\sigma^2)^{-1/2} \exp(-(t-m)^2/2\sigma^2)$ , for all  $t \in \mathbb{R}$ . In this second plug-in step we consider, for the sake of simplicity in our proofs, the nonparametric estimates (25) and (24) instead of (15) and (16). This choice allows us to use similar nonparametric results for both  $\hat{f}_{n,\theta}$  and  $\tilde{I}_{n,\theta}$  (see the proof of Theorem 3.1 ii) and iii)), but the same results should be obtained, at the price of an additional technical lemma, by considering directly the Monte Carlo estimators (15) and (16).

### 3 Identifiability and consistency

#### 3.1 Identifiability

In this section we recall briefly why model (3) is identifiable under conditions similar to those established in [3] and summarized below. Let us define  $\mathcal{F}_s := \{f \in \mathcal{F}; \int_{\mathbb{R}} |x|^s f(x) dx < +\infty\}$  for  $s \geq 1$ , where  $\mathcal{F}$  denotes the set of even pdfs. When  $(f, f_0) \in \mathcal{F}_s$  with  $s \geq 2$ , we denote  $m := \int_{\mathbb{R}} x^2 f(x) dx$  and  $m_0 := \int_{\mathbb{R}} x^2 f_0(x) dx$ .

**Definition 3.1** (*Identifiability*). *Let  $(p_1, \theta_1, f_1, h_1)$  and  $(p_2, \theta_2, f_2, h_2)$  denote two sets of parameters for model (3). The parameter in model (3) is said to be semiparametrically identifiable if*

$$(p_1, \theta_1, f_1(y), h_1(x)) = (p_2, \theta_2, f_2(y), h_2(x)),$$



for  $\lambda^{\otimes 2}$ -almost all  $(x, y) \in \mathbb{R}^2$ , whenever we have :

$$\begin{aligned} & (p_1 f_1(y - \theta_1 \odot x) + (1 - p_1) f_0(y)) h_1(x) \\ &= (p_2 f_2(y - \theta_2 \odot x) + (1 - p_2) f_0(y)) h_2(x), \end{aligned} \quad (26)$$

for  $\lambda^{\otimes 2}$ -almost all  $(x, y) \in \mathbb{R}^2$ .

**Lemma 3.1** *If the Euclidean parameter space  $\Theta$  is a subset of  $\mathbb{R} \times \mathbb{R}^* \setminus \{0, 0\}$ ,  $\text{supp}(f) = \text{supp}(f_0) = \mathbb{R}$ ,  $\text{supp}(h)$  contains at least two intervals respectively in the neighborhood of 0 and  $+\infty$  (or  $-\infty$ ), and the pdfs involved in model (3) satisfy  $(f_0, f) \in \mathcal{F}_3 \times \mathcal{F}_3$ , then the parameter in model (3) is identifiable.*

*Proof.* Integrating (26) with respect to  $y$  over  $\mathbb{R}$ , we then obtain that  $h_1(\cdot) = h_2(\cdot)$   $\lambda$ -almost everywhere. Let  $h(x) := h_1(x)$  for all  $x \in \text{supp}(h) := \text{supp}(h_1) \cap \text{supp}(h_2)$ . Notice now that, for all  $x \in \text{supp}(h)$ , (26) coincides with (5) when considering the generic location parameter  $\mu$  equal to  $\theta \odot x$ . In our case the first three conditional moment equations (given  $\{X = x\}$ ) associated to (26) lead to

$$\left\{ \begin{array}{l} p_1 \theta_1 \odot x = p_2 \theta_2 \odot x, \\ (1 - p_1) m_0 + p_1 ((\theta_1 \odot x)^2 + m_1) = (1 - p_2) m_0 + p_2 ((\theta_2 \odot x)^2 + m_2), \\ p_1 (3((\theta_1 \odot x) m_1 + (\theta_1 \odot x)^3) = p_2 (3((\theta_2 \odot x) m_2 + (\theta_2 \odot x)^3). \end{array} \right. \quad (27)$$

According to [3], the solutions are either, for all  $x \in \text{supp}(h)$ ,  $(p_1, \theta_1 \odot x) = (p_2, \theta_2 \odot x)$ , which implies  $(p_1, \alpha_1, \beta_1) = (p_2, \alpha_2, \beta_2)$ , or

$$\left\{ \begin{array}{lcl} p_2 & = & p_1 \left( \frac{2(\theta_1 \odot x)^2}{3m_1 + (\theta_1 \odot x)^2 - 3m_0} \right), \\ \theta_2 \odot x & = & \theta_1 \odot x + \frac{3m_1 - (\theta_1 \odot x)^2 - 3m_0}{2\theta_1 \odot x}, \\ m_2 & = & m_1 + \frac{(m_1 + (\theta_1 \odot x)^2 - m_0)(3m_1 + (\theta_1 \odot x)^2 - 3m_0)}{4(\theta_1 \odot x)^2}. \end{array} \right. \quad (28)$$

Suppose that  $\beta_1 \neq 0$  and take the limit as  $x \rightarrow +\infty$  in the first row of (28). We then necessarily obtain that  $p_2 = 2p_1$  which is only compatible, when we take the limit as  $x \rightarrow 0$ , with  $m_1 = m_0$ . Hence if  $m_1 \neq m_0$  model (3) is always identifiable. If we suppose  $m_1 = m_0$ , the second row of (28) leads to

$\theta_2 \odot x = (\theta_1 \odot x)/2$ . If we introduce this last relation in the third row of (28) we obtain

$$m_2 - m_1 = \frac{1}{4}(\theta_1 \odot x)^2, \quad x \in \mathbb{R},$$

which is impossible when  $x \rightarrow +\infty$  and thus provides us the global identifiability of model (3).  $\square$

*Remark.* In Lemma 3.1 we have considered for simplicity the case where the slope parameter  $\beta$  is supposed to be different away from zero. Actually this condition can be technically relaxed if we allow  $\theta$  to be equal to  $(\alpha, 0)$  with  $\alpha \neq 0$ . In fact, considering the first row of (28) and taking the limit as  $x \rightarrow +\infty$ , we obtain  $\beta_2 = \beta_1 = 0$ . To conclude, it is then enough to integrate (26) with respect to  $x$  over  $\mathbb{R}$  which leads to discuss the same condition as in [3], p. 735 expression (3). Then Proposition 2 in [3] provides an *almost everywhere*-type identifiability result which unfortunately cannot be strictly compared to the result stated in Lemma 3.1. For this reason we decided to reject  $\theta = (\alpha, 0)$ ,  $\alpha \in \mathbb{R}^*$ , from the sub-parametric space  $\Phi$ , see (6).

## 3.2 Assumptions and statistical complexity

In the following we provide some general conditions that allow us to control the statistical complexity of our model and that insure the validity of basic asymptotic results.

### Regularity conditions (R).

- i) The pdfs  $f$  and  $f_0$  are strictly positive over  $\mathbb{R}$  and belong to  $\mathcal{F}_3$ .
- ii) The pdfs  $f$  and  $f_0$  are twice differentiable over  $\mathbb{R}$  with  $\|f^{(j)}\|_\infty < \infty$  and  $\|f_0^{(j)}\|_\infty < \infty$ , where  $f^{(j)}$  and  $f_0^{(j)}$  denote respectively the  $j$ -th order derivatives of  $f$  and  $f_0$ , for  $j = 1, 2$ .
- iii) The pdf  $h$  satisfies  $\int_{\mathbb{R}} |x|^2 h(x) dx < \infty$ .
- iv) For  $i = 0$  or  $i = 2$ ,

$$\int_{\mathbb{R}^2} |x|^i |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)| h(x) dx dy < \infty,$$

and for  $i = 1$  or  $i = 3$ , and all  $u \in \mathbb{R}$ ,

$$\lim_{y \rightarrow \pm\infty} y^i (F_0(y+u) - F_0(y-u)) = 0.$$

- v) There exist two collections of functions  $\{\ell_{i,j}\}_{0 \leq i \leq j \leq 2}$  and  $\{\ell_{i,j}^0\}_{0 \leq i \leq j \leq 2}$  belonging to  $L_1(\mathbb{R}^2)$  and such that, for all  $(x, y) \in \mathbb{R}^2$  and all  $\theta \in \Theta$

$$|x^i f^{(j)}(y + (\theta - \theta^*) \odot x)| h(x) \leq \ell_{i,j}(x, y),$$

and

$$|x^i f_0^{(j)}(y + \theta \odot x)| h(x) \leq \ell_{i,j}^0(x, y).$$

For all  $z \in \mathbb{C}$ , let  $\bar{z}$  and  $\Im(z)$  the conjugate and imaginary part of  $z$ , respectively. We will also denote  $\bar{f}$ ,  $\bar{f}_0$  the Fourier transforms of  $f$ ,  $f_0$ , and define for all  $\kappa = (\kappa_1, \kappa_2) \in \mathbb{R}^2$ ,  $\nu_\kappa(t) := e^{it\kappa_1} \bar{h}(\kappa_2 t)$ , where  $\bar{h}$  denotes the Fourier transform of  $h$ .

The following conditions mainly insure the contrast property for the function  $d$  defined in (13). We point out that these conditions are not equivalent, as is the case in [5] p. 25, to those established to prove the identifiability property in Lemma 3.1. Loosely speaking the reason of this difference is due to the  $\theta$ -transformation that reduces the Euclidean parameter estimation problem to the analysis of a collection of one-dimensional data, *i.e.* the  $Y_i^\theta$  with  $\theta \in \Phi$ , when the proof of Lemma 3.1 uses strongly the bivariate structure of the original data.

### Contrast condition (C).

- i) The three first moments of  $X$  satisfy  $4E(X)^3 + 3E(X)E(X^2) + E(X^3) \neq 0$ .
- ii) The set of parameters  $\vartheta = (p, \theta) = (p, \alpha, \beta)$  with  $p \neq p_*$  that satisfies

$$p_* \Im(\nu_{\theta-\theta_*}(t)) \bar{f}(t) = (p_* - p) \Im(\nu_\theta(t)) \bar{f}_0(t), \quad t \in \mathbb{R}, \quad (29)$$

is empty or does not belong to the parametric space  $\Theta$ .

- iii) The second order moments of  $f$  and  $f_0$ , respectively denoted  $m$  and  $m_0$ , are supposed to satisfy

$$m \neq m_0 + \frac{\alpha_*^3 + 3\alpha_*^2\beta_*E(X) + 3\alpha_*\beta_*^2E(X^2) + \beta_*^3E(X^3)}{3(\alpha_* + \beta_*E(X))}.$$

*Remark.* Point out that condition C ii), which is necessary to prove that  $d$  is a contrast function over  $\Theta$ , cannot be simplified without more information on  $f$ ,  $f_0$  and  $h$ . We suggest, in the spirit of conditions C1 and C2 in [16], to consider the sufficient and more intuitive *regularity comparison*-type criterion for C ii)

$$\forall \theta \in \Phi, \quad \left| \frac{\Im(\nu_{\theta-\theta_*}(t))}{\Im(\nu_\theta(t))} \frac{\bar{f}(t)}{\bar{f}_0(t)} \right| \longrightarrow +\infty \text{ or } 0, \text{ as } t \rightarrow +\infty, \quad (30)$$

which is valid since, according to (29), the term on left hand side of (30) is equal to  $|p - p_*|/p_* \in [|p - p_*|, 1/\delta]$  which is in contradiction with (30). However condition (29) can directly be discussed in the Gaussian case as done in the appendix, Section 5.1. We prove in particular that there exist sometimes spurious solutions satisfying (29) that have to be removed from the parametric space so they are not detected by our estimation algorithm as shown in Fig. 2.

### Kernel and Bandwidth conditions (K).

- i) The even kernel density function  $K$  is bounded, uniformly continuous, square integrable, of bounded variation and has second order moment.
- ii) The bandwidth  $b_n$  satisfies  $b_n \searrow 0$ ,  $nb_n \rightarrow +\infty$  and  $\sqrt{nb_n^2} = o(1)$ .

**Lemma 3.2** (i) Under conditions (R) the function  $d$  is Lipschitz over  $\Theta$ .

(ii) Under conditions (C) i) and ii) the function  $d$  is a contrast function, i.e. for all  $\vartheta \in \Theta$ ,  $d(\vartheta) \geq 0$  and  $d(\vartheta) = 0$  if and only if  $\vartheta = \vartheta_*$ .

(iii) Under condition (C) iii) we have

$$\ddot{d}(\vartheta_*) = 2 \int_{\mathbb{R}} \dot{H}(y, \vartheta_*) \dot{H}^T(y, \vartheta_*) dQ(y) > 0.$$

(iv) Under conditions (R) and (K), for any  $\gamma > 0$ ,  $d_n$  converges to  $d$  almost surely with the rate

$$\sup_{\vartheta \in \Theta} |d_n(\vartheta) - d(\vartheta)| = o_{a.s.}(n^{-1/2+\gamma}).$$

*Remark .* There exists a simple consistent method to select, in the  $L_1(\mathbb{R}^2)$  sense (recall that our nonparametric consistency results are established for this norm), the best estimator in case of multiple minima of  $d_n$  (which should make suspect that condition (C) is violated). Suppose that, for  $n$  fixed in  $\mathbb{N}^*$ , there exists a finite collection of local minima of  $d_n$ , denoted by  $\hat{\vartheta}_n^{[i]} = (\hat{p}_n^{[i]}, \theta_n^{[i]})$  with  $1 \leq i \leq S < \infty$ . Then we propose to retain a  $\hat{\vartheta}_n$  (in practice unique) satisfying

$$\hat{\vartheta}_n = \hat{\vartheta}_n^{[i_*]}, \quad \text{where } i_* = \arg \min_{1 \leq i \leq S} \left\| \hat{g}_n - \hat{g}_{\hat{\vartheta}_n^{[i]}} \right\|_{L_1},$$

and where for all  $1 \leq i \leq S$ ,  $\hat{g}_{\hat{\vartheta}_n^{[i]}}$  is the plug-in posterior estimator of  $g$  defined by

$$\hat{g}_{\hat{\vartheta}_n^{[i]}} = \hat{p}_n^{[i]} \hat{f}_{\hat{\vartheta}_n^{[i]}} + (1 - \hat{p}_n^{[i]}) f_0, \quad (31)$$

where  $\hat{f}_{\hat{\vartheta}_n^{[i]}}$  corresponds to  $f_n$  defined in (23), when  $\hat{\vartheta}_n = \hat{\vartheta}_n^{[i]}$ . Proceeding in that way, we clearly support the Euclidean parameter estimate that better fit the dataset, this approach being asymptotically consistent as long as the model is identifiable.

PROOF. i) From boundedness and the uniform Lipschitz property of  $H(\cdot, \vartheta)$ , along with the integrability and the integrable Lipschitz property of  $f_\theta(\cdot)$  proved in Sections 5.3, 5.4 and 5.5, there exists a nonnegative constant  $c$  such that for all  $(\vartheta, \vartheta') \in \Theta^2$

$$\begin{aligned} & \left| \int_{\mathbb{R}} H^2(y, \vartheta) dQ(y) - \int_{\mathbb{R}} H^2(y, \vartheta') dQ(y) \right| \\ & \leq \int_{\mathbb{R}} |H(y, \vartheta) + H(y, \vartheta')| |H(y, \vartheta) - H(y, \vartheta')| q(y) dy \\ & \leq c \|\vartheta - \vartheta'\|_3, \end{aligned}$$

which concludes the proof of i).

ii) To clarify the similarity between the semiparametric contamination model (5) studied in [3] and the contaminated regression model (3), we can say that  $f_\theta(\cdot)$  plays the role of  $g(\cdot - \mu)$  and that  $I_\theta(\cdot)$  plays the role of  $f_0(\cdot - \mu)$ .

If  $\vartheta = \vartheta_*$  then  $d(\vartheta) = 0$ . To prove the converse we notice that  $d(\vartheta) = 0$  implies, since  $H_1(\cdot, \vartheta)$  and  $H_2(\cdot, \vartheta)$  are continuous and  $q > 0$  over  $\mathbb{R}$ , that  $H_1(\cdot; \vartheta) = H_2(\cdot; \vartheta)$  which leads, for almost all  $y \in \mathbb{R}$ , to

$$f_\theta(y) - (1 - p)I_\theta(y) = f_\theta(-y) - (1 - p)I_\theta(-y). \quad (32)$$

Using formula (8), we obtain

$$\begin{aligned} p_* \int_{\mathbb{R}} f(y + (\theta - \theta_*) \odot x) h(x) dx + (p - p_*) I_\theta(y) \\ = p_0 \int_{\mathbb{R}} f(-y + (\theta - \theta_*) x) h(x) dx + (p - p_*) I_\theta(-y), \quad y \in \mathbb{R}. \end{aligned}$$

Considering the Fourier transform of the previous equality, using Fubini's Theorem, and noticing that  $\bar{f}$  and  $\bar{f}_0$  are real-valued functions, it follows that

$$\begin{aligned} p_* e^{-it(\alpha - \alpha_*)} \bar{f}(t) \check{h}((\beta - \beta_*)t) + (p - p_*) e^{-it\alpha} \bar{f}_0(t) \check{h}(\beta t) \\ = p_* e^{it(\alpha - \alpha_*)} \bar{f}(t) \bar{h}((\beta - \beta_*)t) + (p - p_*) e^{it\alpha} \bar{f}_0(t) \bar{h}(\beta t), \quad t \in \mathbb{R}. \end{aligned}$$

Using the notation introduced for the writing of condition (C), the previous equation becomes (29).

Suppose that  $p = p_*$  and take the first and third order derivative of (29) at point  $t = 0$ . We then obtain  $\alpha - \alpha_* + (\beta - \beta_*)E(X) = 0$  and

$$\begin{aligned} 3m[\alpha - \alpha_* + (\beta - \beta_*)E(X)] + (\alpha - \alpha_*)^3 + 3(\alpha - \alpha_*)^2(\beta - \beta_*)E(X) \\ + 3(\alpha - \alpha_*)(\beta - \beta_*)^2 E(X^2) + (\beta - \beta_*)^3 E(X^3) = 0, \end{aligned}$$

which naturally leads to

$$(\beta - \beta_*)(4E(X)^3 + 3E(X)E(X^2) + E(X^3)) = 0, \quad (33)$$

and thus implies that  $\theta = \theta_*$  if  $4E(X)^3 + 3E(X)E(X^2) + E(X^3) \neq 0$ .

Suppose now that  $p \neq p^*$ , then condition (C) ii) requires that  $\vartheta = \vartheta_*$ .

iii) First we have

$$\begin{aligned}\ddot{d}(\vartheta_*) &= 2 \int_{\mathbb{R}} \left( \ddot{H}(y, \vartheta_*) H(y, \vartheta_*) + \dot{H}(y, \vartheta_*) \dot{H}^T(y, \vartheta_*) \right) q(y) dy \\ &= 2 \int_{\mathbb{R}} \dot{H}(y, \vartheta_*) \dot{H}^T(y, \vartheta_*) q(y) dy,\end{aligned}\tag{34}$$

according to (9) and the fact that  $H(\cdot, \vartheta_*) = 0$  on  $\mathbb{R}$ . Let  $v$  be a vector in  $\mathbb{R}^3$ . We have

$$v^T \ddot{d}(\vartheta_*) v = 2 \int_{\mathbb{R}} \left( v^T \dot{H}(y, \vartheta_*) \right)^2 q(y) dy \geq 0.\tag{35}$$

It follows that  $\ddot{d}(\vartheta_*)$  is a positive  $3 \times 3$  real valued matrix. Let us show that it is also definite. If  $v \in \mathbb{R}^3$  is a non-null column vector such that  $v^T \ddot{d}(\vartheta_*) v = 0$ , then  $v^T \dot{H}(y, \vartheta_*) = 0$  for almost all  $y \in \mathbb{R}$ . According to (48) in the appendix, we have to discuss the proportionality of  $f$  and  $F_0^{\theta_*}(\cdot) + F_0^{\theta_*}(-\cdot) - 1$ . Because  $f_0$  is an even density, we have from Fubini's theorem

$$f(y) = \frac{\int_{\mathbb{R}} [F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)] h(x) dx}{\int_{\mathbb{R}^2} [F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)] h(x) dx dy}.\tag{36}$$

Using integration by parts and assumption (R) iv), the denominator of the right hand side of (36) can be expressed as follows

$$\begin{aligned}&\int_{\mathbb{R}^2} [F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)] h(x) dx dy \\ &= \int_{\mathbb{R}} \left\{ [y(F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x))]_{-\infty}^{\infty} \right\} h(x) dx \\ &\quad - \int_{\mathbb{R}} \int_{\mathbb{R}} y (f_0(y + \theta_* \odot x) - f_0(y - \theta_* \odot x)) dy h(x) dx \\ &= 2 \int_{\mathbb{R}} (\alpha_* + \beta_* x) h(x) dx = 2(\alpha_* + \beta_* E(X)).\end{aligned}$$

If we calculate now the second-order moment of  $f$  we obtain

$$m := \int_{\mathbb{R}} y^2 f(y) dy = \frac{\int_{\mathbb{R}} x^2 [F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)] h(x) dx}{2(\alpha_* + \beta_* E(X))}.$$

Using integration by parts and assumption (R) iv), the numerator of the right hand-side of (36) can be expressed as follows

$$\begin{aligned}
& \int_{\mathbb{R}^2} y^2 [F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)] h(x) dx dy \\
&= \int_{\mathbb{R}} \left\{ \left[ \frac{y^3}{3} (F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)) \right]_{-\infty}^{\infty} \right\} h(x) dx \\
&\quad - \int_{\mathbb{R}^2} \frac{y^3}{3} (f_0(y + \theta_* \odot x) - f_0(y - \theta_* \odot x)) dy h(x) dx \\
&= 2 \int_{\mathbb{R}^2} \frac{3u^2 \theta \odot x + (\theta \odot x)^3}{3} f_0(u) du h(x) dx \\
&= 2m_0(\alpha_* + \beta_* E(X)) + \frac{2}{3}(\alpha_*^3 + 3\alpha_*^2 \beta_* E(X) + 3\alpha_* \beta_*^2 E(X^2) + \beta_*^3 E(X^3)),
\end{aligned}$$

which leads to a contradiction if (C) iii) is assumed.

iv) This proof, which is a tricky generalization of the proof of Lemma 3.2 iii) given in [5], is relegated to the appendix for convenience, see Section 5.6.  $\square$

**Theorem 3.1** *i) If assumptions (R), (C) and (K) are satisfied then*

$$\|\hat{\vartheta}_n - \vartheta_*\|_3 = o_{a.s.}(n^{-1/4+\gamma}), \quad \gamma > 0.$$

*ii) The estimator  $\hat{f}_n$  of  $f$  defined in (23) converges almost surely in the  $L_1$  sense if  $n^{-1/4+\gamma}/b_n \rightarrow 0$ , for all  $\gamma > 0$ .*

*iii) For any  $\gamma > 0$ , the estimator  $\hat{F}_n$  of  $F$  defined in (22) converges uniformly at the following almost sure rate*

$$\|\hat{F}_n - F\|_{\infty} = O_{a.s.}(n^{-1/4+\gamma}/b_n) + O_{a.s.}(b_n^2), \quad \gamma > 0. \quad (37)$$

*The above rate is optimized by considering  $b_n = n^{-1/12}$ , which choice provides the rate of convergence  $O_{a.s.}(n^{-1/6+\gamma})$ , for all  $\gamma > 0$ .*

*Comment.* Points ii) and iii) reveal the intuitive idea that the bandwidth  $b_n$  must not decrease too fast in order to allow the appropriate positioning of the plug-in-centered data in the expression of  $\hat{\Psi}_{n, \hat{\theta}_n}$ . In fact the  $Y_i^{\hat{\theta}_n}$  need to be sufficiently close to the  $Y_i^{\theta_*}$ , and  $b_n$  not too small (the smaller  $b_n$  is the more



we “freeze” the kernel estimator), if we want a good agreement between  $\hat{\Psi}_{n,\hat{\theta}_n}$  and  $\hat{\Psi}_{n,\theta_*}$  which are known to converge to the true  $\Psi_{\theta_*}$  involved in expression (9).

PROOF. i) The proof follows entirely the proof of Theorem 3.1 in [5] and uses the technical results proved in Lemma 3.2.

ii) Consider the following decomposition:

$$\begin{aligned}
|\hat{f}_n - f| &= \left| \left[ \frac{1}{\hat{p}_n} \hat{\Psi}_{n,\hat{\theta}_n} - \frac{1}{p_*} \Psi_{\theta_*} \right] + \left[ \frac{1 - \hat{p}_n}{\hat{p}_n} \tilde{I}_{n,\hat{\theta}_n} - \frac{1 - p_*}{p_*} I_{\theta_*} \right] \right| \\
&= \left| \left[ \frac{1}{\hat{p}_n} (\hat{\Psi}_{n,\hat{\theta}_n} - \hat{\Psi}_{n,\theta_*}) + \frac{1}{\hat{p}_n} \hat{\Psi}_{n,\theta_*} - \frac{1}{p_*} \Psi_{\theta_*} \right] \right. \\
&\quad \left. + \left[ \frac{1 - \hat{p}_n}{\hat{p}_n} (\tilde{I}_{n,\hat{\theta}_n} - \tilde{I}_{n,\theta_*}) + \frac{1 - \hat{p}_n}{\hat{p}_n} \tilde{I}_{n,\theta_*} - I_{\theta_*} \right] \right| \\
&\leq \frac{1}{\hat{p}_n} \left( |\hat{\Psi}_{n,\hat{\theta}_n} - \hat{\Psi}_{n,\theta_*}| + |\hat{\Psi}_{n,\theta_*} - \Psi_{\theta_*}| \right) + \Psi_{\theta_*} \left| \frac{1}{\hat{p}_n} - \frac{1}{p_*} \right| \\
&\quad + \frac{1 - \hat{p}_n}{\hat{p}_n} \left( |\tilde{I}_{n,\hat{\theta}_n} - \tilde{I}_{n,\theta_*}| + |\tilde{I}_{n,\theta_*} - I_{\theta_*}| \right) \\
&\quad + I_{\theta_*} \left| \frac{1 - \hat{p}_n}{\hat{p}_n} - \frac{1 - p_*}{p_*} \right|. \tag{38}
\end{aligned}$$

It is now enough to study the behavior of  $|\hat{\Psi}_{n,\hat{\theta}_n} - \hat{\Psi}_{n,\theta_*}|$  and  $|\tilde{I}_{n,\hat{\theta}_n} - \tilde{I}_{n,\theta_*}|$ . For all  $t \in \mathbb{R}$ , we have

$$|\hat{\Psi}_{n,\hat{\theta}_n}(t) - \hat{\Psi}_{n,\theta_*}(t)| \leq \frac{1}{nb_n} \sum_{k=1}^n \left| K\left(\frac{t - Y_i^{\hat{\theta}_n}}{b_n}\right) - K\left(\frac{t - Y_i^{\theta_*}}{b_n}\right) \right|. \tag{39}$$

Consider  $K$  a centered normalized gaussian kernel. We propose to study in a generic way the difference of kernels involved in the right hand side of the above expression. For all  $(w, z) \in \mathbb{R}^2$ , and letting  $h := (z - w)/b$ , we write the second-order Taylor expansion with integral remaining term:

$$K\left(\frac{t - w}{b}\right) - K\left(\frac{t - z}{b}\right) = h \dot{K}\left(\frac{t - z}{b}\right) + \frac{h^2}{2} \int_0^1 (1 - u) \ddot{K}\left(\frac{t - m_u}{b}\right) du$$

where  $m_u := (1 - u)z + uw$ . Noticing that

$$\dot{K}\left(\frac{t - z}{b}\right) = (t - z) \mathcal{N}_{z,b^2}(t), \quad \text{and} \quad \ddot{K}\left(\frac{t - m_u}{b}\right) = \left(-b + \frac{(t - m_u)^2}{b}\right) \mathcal{N}_{m_u,b^2}(t) dt,$$

it thus follows that

$$\begin{aligned}
& \frac{1}{b} \int_{\mathbb{R}} \left| K\left(\frac{t-w}{b}\right) - K\left(\frac{t-z}{b}\right) \right| dt \\
& \leq h \int_{\mathbb{R}} |t-z| \frac{\mathcal{N}_{z,b^2}(t)}{b} dt + \frac{h^2}{2} \int_{\mathbb{R}} \left(1 + \frac{(t-m_u)^2}{b^2}\right) \mathcal{N}_{m_u,b^2}(t) dt \\
& \leq -h \sqrt{\frac{2}{\pi}} \int_0^\infty -\frac{t}{b^2} \exp\left(-\frac{t^2}{2b^2}\right) dt + h^2 \\
& \leq \sqrt{\frac{2}{\pi}} h + h^2.
\end{aligned} \tag{40}$$

Replacing  $w, z$  respectively by the  $Y_i^{\theta_n}$  and  $Y_i^{\theta_*}$ , and  $b$  by  $b_n$  in (40) we then obtain from (39) the following bound for the  $L_1$  error:

$$\|\hat{\Psi}_{n,\hat{\theta}_n}(t) - \hat{\Psi}_{n,\theta_*}(t)\|_{L_1} \leq \frac{C\|\theta_n - \theta_*\|_2}{b_n} \times \frac{1}{n} \sum_{k=1}^n (|X_i| + |X_i|^2), \tag{41}$$

the same kind of bound being available for  $\|\tilde{I}_{n,\hat{\theta}_n}(t) - \tilde{I}_{n,\theta_*}(t)\|_{L_1}$ . In conclusion, according to the decomposition (38), point i) of Theorem 3.1, the respective  $L_1$  *a.s.* convergence of  $\hat{\Psi}_{n,\theta_*}$  and  $\tilde{I}_{n,\theta_*}$  towards  $\Psi_{\theta_*}$  and  $I_{\theta_*}$  under (20), we get from (41) and the strong law of large numbers that  $\|\hat{f}_n - f\|_{L_1} \rightarrow 0$  almost surely as  $n \rightarrow \infty$  whenever  $n^{-1/4+\gamma}/b_n = o(1)$ .

iii) The proof uses an integrated version of decomposition (38) and the fact that, for all  $y \in \mathbb{R}$ , the approximation  $|\hat{F}_{n,\hat{\theta}_n} - \hat{F}_{n,\theta_*}|(y)$  is controlled by

$$\begin{aligned}
|\hat{F}_{n,\hat{\theta}_n} - \hat{F}_{n,\theta_*}|(y) &= \left| \int_{-\infty}^y \frac{1}{n} \sum_{i=1}^n K\left(\frac{t - Y_i^{\hat{\theta}_n}}{b_n}\right) - K\left(\frac{t - Y_i^{\theta_*}}{b_n}\right) dt \right| \\
&\leq \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}} \left| K\left(\frac{t - Y_i^{\hat{\theta}_n}}{b_n}\right) - K\left(\frac{t - Y_i^{\theta_*}}{b_n}\right) \right| dt \\
&\leq \frac{C\|\theta_n - \theta_*\|_2}{b_n} \times \frac{1}{n} \sum_{k=1}^n (|X_i| + |X_i|^2),
\end{aligned} \tag{42}$$

the last term in the right hand side of above inequality being independent from  $y$ . The same bound holds for  $|\tilde{I}_{n,\hat{\theta}_n} - \tilde{I}_{n,\theta_*}|(y)$  by an identical argument. To conclude, it is enough to use (42) and Corollary 1 p. 766 in [22] which allows us to control the terms  $\|\hat{F}_{n,\theta_*} - F_{\theta_*}\|_\infty$  and  $\|\tilde{I}_{n,\theta_*} - I_{\theta_*}\|_\infty$ , to obtain (37). The

rate on right hand side of (37) is optimized by considering  $b_n = n^{-1/12}$  which then turns into  $O_{a.s.}(n^{-1/6+\gamma})$ , for all  $\gamma > 0$ .

□

## 4 Numerical experiments

### 4.1 Role of the $\theta$ -transformation

We propose in this section to highlight the role played by the  $\theta$ -transformation, see (7), in our method. For this purpose, we consider an example which corresponds to model (5) taking  $p_* = 0.7$ ,  $\alpha_* = 2$ ,  $\beta_* = 1$ ,  $\varepsilon_1^{[j]} \sim \mathcal{N}(0, 1)$ ,  $j = 0, 1$  and  $X_1 \sim \mathcal{N}(2, 3)$ . In Fig. 1 we plot successively a simulated data set  $(X_i, Y_i)_{1 \leq i \leq n}$ , corresponding to the previous description with  $n = 200$ , and the two  $\theta$ -transformed datasets obtained with  $\theta = (1, 0.5)$  and  $\theta = \theta^* = (2, 1)$ . These figures are completed by adding their corresponding 2nd-coordinate sam-

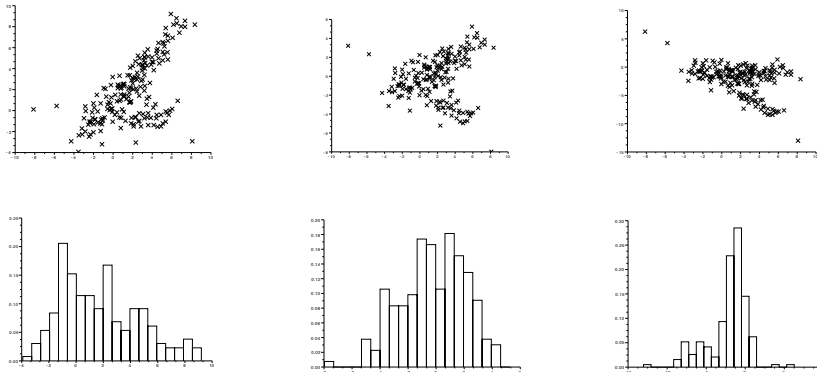


Figure 1: First row: resp. plot of an original data  $(X_i, Y_i)_{1 \leq i \leq n}$  according to model (3) with  $n = 200$ , plot of a wrong  $\theta$ -transformation ( $\theta \neq \theta^*$ ), plot of the true  $\theta^*$ -transformation. Second row: resp. histograms of the corresponding first row 2nd-coordinate sample data.

ple data histograms. Note that these histograms are empirical estimates of the densities  $f_\theta$ , by formula (8), with  $\theta$  respectively equal to  $(0, 0)$ ,  $(1, 0.5)$  and  $(2, 1)$ . We see clearly through these three situations how a progressive transformation of the data allows one to reach a tractable situation in the sense that

it looks strongly like the semiparametric contamination model (5) studied in [3] and [5] where a known density is mixed with a symmetric unknown density, which corresponds to the behavior observed in the second row, third column histogram in Fig. 1. Loosely speaking the second idea of our method consists in arguing that once  $\theta$  is close to  $\theta_*$  we are allowed to estimate the proportion  $p$  according to a [5] type-method which corresponds to the minimization step (21). In contrast to this technically satisfying idea, the  $\theta$ -transformation and the choice of the weight distribution  $Q$  introduced in (13) are two sources of serious difficulties. In fact when  $\beta_*$  is large and the law of the design data has heavy tails with respect to the tails of  $f$ , then the  $\theta_*$  transformation will move the points coming from the  $F_0$ -population and located far from the origin, to extremely distant positions, which implies intuitively that the integral type density involved in (7) should be extremely heavily tailed. Thus in order to capture the information contained in the tails of the  $\theta$ -transformed data set it is important to weight sufficiently the empirical index of symmetry  $H^2(x; p, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta})$  of expression (14) for large values of  $x$ , which reduces to choosing an instrumental distribution  $Q$  with non-negligible tails with respect to  $F_{\theta_*}$ .

## 4.2 Otimization procedure and simulation study

The aim of this section is to illustrate graphically, on a two-dimensionnal examples, the behavior of the empirical distance  $d_n(p, 0, \beta)$  (the parameter  $\alpha$  is assumed to be equal to zero) when  $p$  and  $\beta$  lie close to the true value of the parameter. For simplicity the parameter will still be denoted  $\vartheta := (p, \beta)$ , with  $\theta := \beta$  and  $d_n(\vartheta) := d_n(p, 0, \beta)$ . The interest of this study is to understand closely the influence of the mixing proportion  $p$  and the regression coefficient  $\beta$  on the shape of the contrast function  $d$  (flatness, sharpness, smoothness, etc.). Our models are denoted M1 and M2 and defined according to (2) as follows

**M1:**  $p_* = 0.7$ ,  $\beta_* = 1$ ,  $V \sim Q = \mathcal{N}(0, 4^2)$ ,  $\varepsilon^{[j]} \sim \mathcal{N}(0, 1)$ ,  $X \sim \mathcal{N}(0, 3^2)$ ,

**M2:**  $p_* = 0.3$ ,  $\beta_* = 1$ ,  $V \sim Q = \mathcal{N}(0, 2^2)$ ,  $\varepsilon^{[j]} \sim \mathcal{N}(0, 1)$ ,  $X \sim \mathcal{N}(0, 3^2)$ ,

where  $j = 0, 1$ .

In Fig. 2 we plot the mapping  $(p, \beta) \mapsto d_n(p, \beta)$  obtained from an M1-sample, resp. M2-sample, of size  $n = 100$ , where  $(p, \beta) \in \Theta_1 = [0.5, 0.8] \times [0.9, 1.1]$ , resp.  $(p, \beta) \in \Theta_2 = [0.1, 0.6] \times [0.6, 1.4]$ . Notice that according to discussion (CG) at the end of Section 5.1, model M2 is not necessarily consistently estimated if the parameter space  $\Theta_2$  contains the spurious solution  $\vartheta_{**} = (2p_*, \beta_*/2)$ , which is voluntary the case here. In practice, Fig. 2 is

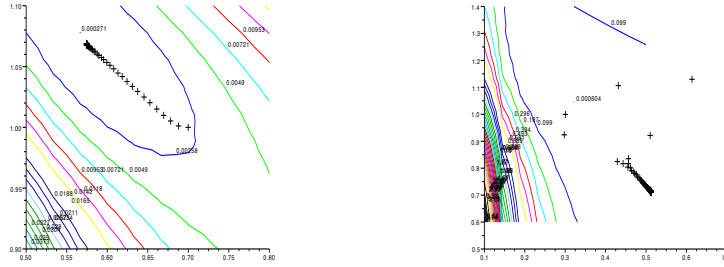


Figure 2: Plot of  $(p, \beta) \mapsto d_n(p, 0, \beta)$  with  $n = 100$ ,  $\beta_* = 1$ ,  $\varepsilon_1^{[j]} \sim \mathcal{N}(0, 1)$ ,  $j = 0, 1$ ,  $X_1 \sim \mathcal{N}(2, 3^2)$ , with the difference that on the left hand side  $p_* = 0.7$ ,  $V_1 \sim \mathcal{N}(0, 4^2)$ , when on the right hand side  $p_* = 0.3$   $V_1 \sim \mathcal{N}(0, 2^2)$ .

obtained using the Scilab `contour2d` function which plots the level curves of  $d_n$  evaluated on a homogeneous  $10 \times 10$  grid of the rectangular domain  $[0.5, 0.8] \times [0.9, 1.1]$ . Fig. 2 shows that the graph of  $d_n$  looks like a sharp valley with a flat trough when  $\beta$  is located near  $\beta_*$  and  $p$  ranges  $[0.5, 0.8]$ . Even if on this simulated example the argmin of  $d_n$  is very close to the true value of the parameter, the previous remark suggests that the estimation of the mixing proportion will be less robust than the estimation of the regression coefficient. The observation of the second plot in Fig. 2 is more unexpected since the graph of  $d_n$  does not really look like a contrast function with its high near  $p = 0.1$  and its very large and flat trough that covers most of  $\Theta_2$  suggesting a strong lack of robustness of our estimating method in that kind of situation. To validate these thoughts we propose to apply a large sample study on the

example and a third intermediary one obtained by considering  $p_* = 0.3$  and  $V \sim \mathcal{N}(0, 4^2)$ . The results of this study will be summarized in Table 1. First we present the numerical approach used to approximate our M-estimator (21).

*Gradient algorithm and tuning parameters.* The gradient optimization procedure (programmed with Scilab) used to compute our M-estimator  $\hat{\vartheta}_n = (\hat{p}_n, \hat{\beta}_n)$  is defined as follows:

- (i) **Initialization:**  $\vartheta_1 = \vartheta_*, \vartheta_2 = \vartheta_* + \delta$ ;
- (ii) **while**  $\|\vartheta_2 - \vartheta_1\|_2 > \epsilon$  **do**  $\vartheta_1 = \vartheta_2$  **and**  $\vartheta_2 = \vartheta_1 - \gamma^T \dot{d}_n(\vartheta_1)$ ;
- (iii) **else**  $\hat{\vartheta}_n = \vartheta_2$ ,

where  $\delta \in \mathbb{R}^2$  is used to create a small perturbation of the initial value,  $\epsilon > 0$  defines the wanted stabilization level in the stopping algorithm procedure, and  $\gamma \in \mathbb{R}^{+*2}$  is a scale parameter that needs to be hand-tuned for good efficiency in practice (to avoid reverberation phenomena when the score function becomes abruptly sharp). The score function  $\dot{d}_n := \left( \frac{\partial}{\partial p} d_n, \frac{\partial}{\partial \beta} d_n \right)^T$  can be expressed into a closed form, *i.e.*

$$\begin{aligned} \frac{\partial}{\partial p} d_n(\vartheta) &= 2 \int_{\mathbb{R}} h_{n,p}(y, \vartheta) H_n(y, \vartheta) dQ_n(y), \\ \frac{\partial}{\partial \beta} d_n(\vartheta) &= 2 \int_{\mathbb{R}} h_{n,\beta}(y, \vartheta) H_n(y, \vartheta) dQ_n(y), \end{aligned}$$

where for all  $y \in \mathbb{R}$ ,

$$\begin{aligned} h_{n,p}(y, \vartheta) &:= \frac{\partial}{\partial p} H_n(y, \vartheta) = -\frac{1}{p^2} \left( \tilde{F}_{n,\beta}(y) + \tilde{F}_{n,\beta}(-y) - [\hat{J}_{n,\beta}(y) + \hat{J}_{n,\beta}(-y)] \right) \\ h_{n,\beta}(y, \vartheta) &:= \frac{\partial}{\partial \beta} H_n(y, \vartheta) = \frac{1}{p} \left( \tilde{\Psi}_{n,\beta}(y) + \tilde{\Psi}_{n,\beta}(-y) \right) - \frac{1-p}{p} (j_{n,\beta}(y) + j_{n,\beta}(-y)), \end{aligned}$$

and where, from (18) and (19)

$$\begin{aligned}
\tilde{\Psi}_{n,\beta}(y) &:= \frac{\partial}{\partial \beta} \tilde{F}_{n,\beta}(y) \\
&= \frac{1}{nb_n} \sum_{i=1}^n \frac{\partial}{\partial \beta} \int_{-\infty}^y K\left(\frac{t - (Y_i - \beta X_i)}{b_n}\right) dt \\
&= \frac{1}{n} \sum_{i=1}^n \frac{\partial}{\partial \beta} \int_{-\infty}^{\frac{y + \beta X_i - Y_i}{b_n}} K(u) du \\
&= \frac{1}{nb_n} \sum_{i=1}^n X_i K\left(\frac{y + \beta X_i - Y_i}{b_n}\right),
\end{aligned}$$

and similarly, from (17)

$$\begin{aligned}
j_{n,\beta}(y) &:= \frac{\partial}{\partial \beta} \hat{J}_{n,\beta}(y) \\
&= \frac{1}{n} \sum_{i=1}^n X_i f_0(y + \beta X_i).
\end{aligned}$$

The kernel  $K$  used to compute (19), is a triangular kernel defined by

$$K(x) = (1 - |x|) \mathbf{1}_{-1 \leq x \leq 1}, \quad x \in \mathbb{R},$$

and the bandwidth  $b_n = \sqrt{1 + 4p(1 - p)}(4/(3n))^{1/5}$  (proposed by [7] for gaussian distributions and implemented in  $\mathbf{R}$ ), both obviously satisfying condition (K). The results summarized in Table 1 were obtained with the following hand-tuned parameters:  $\delta = 0.01$ ,  $\epsilon = 0.005$ ,  $\gamma = (0.2, 0.5)^T$ , and an example of stabilization for this set of tuning parameters is illustrated in Fig 2, where the successive positions (until stabilization) of our algorithm are depicted by cross symbols.

*Comments on Table 1.* First of all, it is interesting to compare the performances summarized in rows 1–3 of Table 1 to those obtained in [5], p. 35, Table 1 where the model of interest is (5), with  $p = 0.7$ ,  $\mu = 3$ , and  $f_0$  and  $f$  are respectively the pdfs corresponding to the  $\mathcal{N}(0, 1)$  and  $\mathcal{N}(0, (1/2)^2)$  distributions. Even if these two models are not strictly comparable we think that it is interesting, in order to highlight the drawbacks induced by the  $\theta$ -transformation and the choice of  $Q$  discussed above, to compare pairwise the performance obtained on

Table 1: Mean and Std. Dev. of 100 estimates of  $p, \beta$ .

$n$	$(p_*, \beta_*, \sigma_V)$	Empirical means	Standard deviation
100	(0.7,1,4)	(0.7055,1.0051)	(0.0373,0.0697)
200	(0.7,1,4)	(0.6976,0.9965)	(0.0307,0.0590)
500	(0.7,1,4)	(0.6954,1.0059)	(0.0296,0.0358)
100	(0.3,1,4)	(0.3100,0.9581)	(0.0577,0.1252)
200	(0.3,1,4)	(0.2965,0.9851)	(0.0501,0.0855)
500	(0.3,1,4)	(0.2975,1.0178)	(0.0284,0.0414)
100	(0.3,1,2)	(0.3971, 0.8587)	(0.0942, 0.2213)
200	(0.3,1,2)	(0.3982,0.9149)	(0.0835,0.1900)
500	(0.3,1,2)	(0.3315, 0.9683)	(0.0524, 0.1067)

the mixing proportion  $p$  and the parameters that influence the location of the  $F$ -population, *i.e.*  $\beta$  and  $\mu$ . From the numerical point of view, we easily check that the bias of our estimators, for both models, is negligible. However it also appears that the standard deviation associated to  $(\hat{p}_n, \hat{\beta}_n)$  decreases significantly slower than the standard deviation associated to  $(\hat{p}_n, \hat{\mu}_n)$  when  $n$  grows. The performance summarized in rows 4–6 of Table 1, which corresponds to  $p = 0.3$  (and hence signifies that the population that will move far from its original position due to the  $\theta$ -transformation will be more important), is very instructive. We observe that for small  $n$  ( $n = 100, 200$ ) the standard deviations associated to  $(\hat{p}_n, \hat{\beta}_n)$  are dramatically large compared to those obtained with  $p = 0.7$ . Let  $\text{std}(n, p_*, \beta_*, \sigma_V)$  the couple of standard deviations calculated in the last column of Table 1 under  $(n, p_*, \beta_*, \sigma_V)$ . If we compute componentwise the ratios  $\text{std}(n, 0.3, 1, 4)/\text{std}(n, 0.7, 1, 4)$  respectively for  $n = 100, 200, 500$  we obtain approximately (1.54, 1.8), (1.67, 1.44), and (0.95, 1.17) which seems to suggest that when  $n$  becomes large the side effect of the  $\theta$ -transformation vanishes (probably thanks to the size of  $n$ , which increases globally the precision of the empirical estimates, and the tails of  $Q$ , that allow the algorithm to take these improvements into account efficiently). The performances summarized in rows 7–9 of Table 1, seems to confirm the concerns expressed about model M2. We recall that model M2 is badly affected by the two following draw-



backs : smallness of  $p_*$  (synonymous with important population shifted far by the  $\theta$ -transformation and existence of a spurious solution) and a smallness of  $\sigma_V$  which is then clearly not sufficient to counteract the smallness of  $p_*$  (and its consequences). We think in particular that, in model M2, the empirical contrast  $d_n$  is more easily closer to 0 under  $\beta_{**} = \beta_*/2$  since as explained in Section 4.1., this value is then significantly smaller than  $\beta_*$ . This last remark explains why, in spite of the fact that our algorithms were initialized at the true parameter value, our estimates are strongly biased (attracted quite often by the spurious solution  $\vartheta_{**}$ ).

*Robustness with respect to the symmetry assumption.* We propose to conclude this simulation study by testing our method in situations where the law of the error  $\varepsilon^{[1]}$  is no longer symmetric. For this purpose we consider again model M1 and replace the distribution of  $\varepsilon^{[1]}$  by the mixture

$$\lambda\mathcal{N}(-0.7, 1/\sqrt{2}) + (1 - \lambda)\mathcal{N}(0.7\lambda/(1 - \lambda), 1/\sqrt{2})$$

which pdf, denoted  $f_\lambda$ , is nonsymmetric if  $\lambda \neq 0.5$  but has a mean equal to 0 and a variance equal to 0.5 for all  $\lambda \in (0, 1)$ . In our simulations we consider successively  $\lambda = 0.5, 0.55, 0.6$  which leads to consider pdfs for  $\varepsilon^{[1]}$  which graphs are plotted in Fig 3. Some performances of our method on these examples are

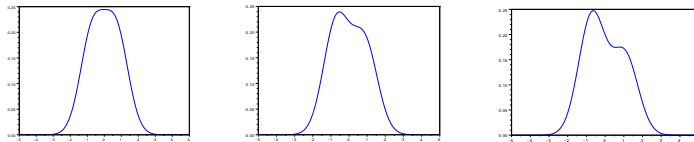


Figure 3: Graphs of the pdfs corresponding to the mixture distribution  $\lambda\mathcal{N}(-0.7, 1/\sqrt{2}) + (1 - \lambda)\mathcal{N}(0.7\lambda/(1 - \lambda), 1/\sqrt{2})$ , obtained by considering  $\lambda = 0.5, 0.55, 0.6$ .

summarized in Table 2.

*Comments on Table 2.* Note that when  $\lambda = 0.5$  (symmetric case) the performances of our method are very close to those obtain on model M1. However

Table 2: Mean and Std. Dev. of 100 estimates of  $p$ ,  $\beta$ .

$n$	$\lambda$	Empirical means	Standard deviation
100	0.5	(0.7035,1.0229)	(0.0427,0.0814)
200	0.5	(0.7012,1.0068)	(0.0390,0.0774)
500	0.5	(0.6997,1.0059)	(0.0244,0.0488)
100	0.55	(0.6854,1.0837)	(0.0485,0.0858)
200	0.55	(0.6890,1.0805)	(0.0431,0.0716)
500	0.55	(0.6922,1.0699)	(0.0377,0.0519)
100	0.6	(0.6731,1.1314)	(0.0543,0.0952)
200	0.6	(0.6693,1.1061)	(0.0490,0.0868)
500	0.6	(0.6775,1.0928)	(0.0392,0.0557)

for  $n = 100, 200$  the standard deviation of our estimates is larger than those obtained in the M1 model, when for  $n = 500$  the standard deviation becomes slightly smaller. This behavior can probably be explained by the fact that the graph of  $f_{0.5}$  is flat on its top which intuitively do not help much in locating the axis of symmetry for small values of  $n$ . On the other hand we can expect that for  $n = 500$ , helped with the fact that  $\text{var}(\varepsilon^{[1]})$  is here equal to 0.5 when it was equal to 1 in M1, our nonparametric estimators perform better than in model M1 which should explain the good performances observe in the third row of Table 2. When  $\lambda = 0.55, 0.6$  it appears that the parameter  $\beta$  is always overestimated. This phenomenon can be explained by the fact that our method try to determine a pseudo-axis of symmetry adapted to the shapeless graph of  $f_\lambda$  which qualitatively is placed on the left side of the origin. This remark implies that the  $\theta$ -transformation needed to transform the first integral in (8) into an almost even density (see Fig. 2) have to contain a  $\beta$  greater than  $\beta_*$ .

## 5 Appendix

### 5.1 Conditions (R) and (C) in the Gaussian Case

In this section we discuss conditions (R) and (C) when the true underlying model is a contaminated Gaussian regression model with Gaussian design, *i.e.*,

$f, f_0, h$  are respectively the pdfs of the  $\mathcal{N}(0, m), \mathcal{N}(0, m_0)$ , and  $\mathcal{N}(E(X), \sigma_h^2)$  distributions.

*Comments on condition (R).* Conditions (R) i-iii) are standard and easy to verify in the above model. On the other hand, it is interesting to show how conditions (R) iv-v) arise naturally in this case.

*Condition (R) iv).* We show for simplicity that the first condition in (R) iv) (the same kind of proof works also for the second one) holds when  $i = 0$ ,  $\theta^* = (\alpha_*, \beta_*) \in \mathbb{R}^{+*2}$  and  $m_0 = 1$ . We write the decomposition

$$\begin{aligned} & |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)| = \\ & |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)| \mathbf{1}_{y > 1 - \theta_* \odot x, x < -\frac{\alpha_*}{\beta_*}} \\ & + |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)| \mathbf{1}_{y > 1 + \theta_* \odot x, x \geq -\frac{\alpha_*}{\beta_*}} \\ & + |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)| \mathbf{1}_{y < -1 + \theta_* \odot x, x < -\frac{\alpha_*}{\beta_*}} \\ & + |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)| \mathbf{1}_{y < -1 - \theta_* \odot x, x \geq -\frac{\alpha_*}{\beta_*}} \\ & + |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)| \mathbf{1}_{-1 + \theta_* \odot x \leq y \leq 1 - \theta_* \odot x, x < -\frac{\alpha_*}{\beta_*}} \\ & + |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)| \mathbf{1}_{-1 - \theta_* \odot x \leq y \leq 1 + \theta_* \odot x, x \geq -\frac{\alpha_*}{\beta_*}}. \end{aligned}$$

Consider the first term on the right hand side of the above decomposition (the three following terms being treated in entirely same way). For all  $y > 1 - \theta_* \odot x$  with  $x < -\alpha_*/\beta_*$  we have  $y - \theta_* \odot x > y + \theta_* \odot x > 1$ . Since for  $t > 0$  large enough, the inequality (43) is valid

$$\frac{\exp(-t^2)}{\sqrt{2\pi}} \left( \frac{1}{t} - \frac{1}{t^3} \right) \leq 1 - F_0(t) \leq \frac{\exp(-t^2)}{t\sqrt{2\pi}}, \quad (43)$$

we have in particular that for all  $t > 1$ ,  $0 \leq 1 - F_0(t) \leq \exp(-t^2)/\sqrt{2\pi}$ . Hence it follows that for  $y > 1 - \theta_* \odot x$  with  $x < -\alpha_*/\beta_*$ :

$$|F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)| \leq \frac{\exp(-(y + \theta_* \odot x)^2) + \exp(-(y - \theta_* \odot x)^2)}{\sqrt{2\pi}},$$

which proves that this first term is  $h(x)dxdy$  integrable. Let us now sum the

last two terms of the above decomposition and notice that

$$\begin{aligned}
& |F_0(y + \theta_* \odot x) - F_0(y - \theta_* \odot x)| \\
& \times \left( \mathbf{1}_{-1+\theta_* \odot x \leq y \leq 1-\theta_* \odot x, x < -\frac{\alpha}{\beta_*}} + \mathbf{1}_{-1-\theta_* \odot x \leq y \leq 1+\theta_* \odot x, x \geq -\frac{\alpha}{\beta_*}} \right) \\
& \leq 2 \mathbf{1}_{-1-|\theta_* \odot x| \leq y \leq 1+|\theta_* \odot x|}.
\end{aligned}$$

We thus prove that this sum of terms is also  $h(x)dxdy$  integrable.

*Condition (R) v).* We consider for simplicity the construction of the bounding function  $\ell_{1,1}^0$  when  $(\alpha, \beta) \in \Phi = [\underline{\alpha}, \bar{\alpha}] \times [\underline{\beta}, \bar{\beta}]$ , with  $(\underline{\alpha}, \underline{\beta}) \in \mathbb{R}^{+*2}$  and  $m = m_0 = 1$ . Notice first that for all  $(x, y) \in \mathbb{R}^2$

$$|f_0^{(1)}(y + \theta \odot x)| \leq \frac{|y| + \bar{\alpha} + \bar{\beta}|x|}{\sqrt{2\pi}} \exp\left(-\frac{(y + \alpha + \beta x)^2}{2}\right).$$

Secondly it is easy to check that for all  $(x, y) \in \mathbb{R}^2$  and all  $\theta \in \Phi$ :

$$\begin{aligned}
& \exp\left(-\frac{(y + \alpha + \beta x)^2}{2}\right) \\
& \leq \exp\left(-\frac{(y + \underline{\beta}x)^2}{2}\right) \mathbf{1}_{x \geq 0, y \geq 0} \\
& + \exp\left(-\frac{\min(|y + \underline{\alpha} + \underline{\beta}x|, |y + \bar{\alpha} + \bar{\beta}x|)^2}{2}\right) \mathbf{1}_{x \geq 0, y \leq 0} \\
& + \exp\left(-\frac{\min(|y + \underline{\alpha} + \bar{\beta}x|, |y + \bar{\alpha} + \underline{\beta}x|)^2}{2}\right) \mathbf{1}_{x \leq 0, y \in \mathbb{R}} \\
& \leq B_\Phi(x, y),
\end{aligned}$$

where

$$\begin{aligned}
B_\Phi(x, y) &:= \exp\left(-\frac{(y + \underline{\beta}x)^2}{2}\right) + \exp\left(-\frac{(y + \underline{\alpha} + \underline{\beta}x)^2}{2}\right) \\
&+ \exp\left(-\frac{(y + \bar{\alpha} + \bar{\beta}x)^2}{2}\right) + \exp\left(-\frac{(y + \underline{\alpha} + \bar{\beta}x)^2}{2}\right) \\
&+ \exp\left(-\frac{(y + \bar{\alpha} + \underline{\beta}x)^2}{2}\right).
\end{aligned}$$

Thus we can propose  $\ell_{1,1}^0(x, y) = |x|(|y| + \bar{\alpha} + \bar{\beta}|x|)/\sqrt{2\pi}B_\Phi(x, y)\exp(-x^2/2)$ , which clearly belongs to  $L_1(\mathbb{R}^2)$ , as a candidate for the uniformly bounding

function satisfying condition (R) v).

*Comments on condition (C).* In the whole Gaussian case, expression of (29) becomes:

$$\begin{aligned} & p_* \sin((\alpha - \alpha_*) + (\beta - \beta_*)E(X))t \exp\left(-\frac{t^2}{2}(\sigma_h^2(\beta - \beta_*)^2 + m)\right) \\ &= (p_* - p) \sin((\alpha + \beta E(X))t) \exp\left(-\frac{t^2}{2}(\sigma_h^2\beta^2 + m_0)\right). \end{aligned} \quad (44)$$

We suppose first that  $p \neq p_*$ , and denote  $\xi := \alpha + \beta E(X)$ ,  $\xi_* := \alpha_* + \beta_* E(X)$ ,  $\Sigma_{\beta-\beta_*} := \sigma_h^2(\beta - \beta_*)^2 + m$ , and  $\Sigma_\beta := \sigma_h^2\beta^2 + m_0$ . Taking the first and third order derivative of (44) at point  $t = 0$  we get the conditions

$$p_*\xi_* = p\xi, \quad \text{and} \quad (p - p_*)\xi(3\Sigma_{\beta-\beta_*} + \xi^2) + p_*(\xi - \xi_*)(3\Sigma_\beta + (\xi - \xi_*)^2) = 0. \quad (45)$$

Introducing the first relation in (45) into the second one, we obtain

$$\frac{p - p_*}{p} \xi_* p_* \left( 3[\Sigma_\beta - \Sigma_{\beta-\beta_*}] + \frac{2p_*p - p^2}{p^2} \xi_*^2 \right) = 0. \quad (46)$$

Now we observe that, to insure the validity of expression (44), the factors multiplied by the sin terms on both sides of (44) must be, at least, equivalent as  $t \rightarrow \infty$ . This last remark implies that  $\Sigma_\beta = \Sigma_{\beta-\beta_*}$ , or equivalently  $\beta = \beta_*/2 + (m_0 - m)/2\beta_*\sigma_h^2$ , and thus (46) leads to

$$p = 2p_*. \quad (47)$$

Using now the first relation in (44) and (47), we then obtain  $\alpha = \alpha_*/2 + E(X)(m_0 - m)/2\beta_*\sigma_h^2$ .

The consequences of the previous comments can be presented as follows:

*Discussion (CG):*

- i) If  $p_* > 1/2$  then the set of parameters  $\vartheta \in \Theta$  satisfying condition (44) is always empty, since  $p = 2p_* > 1$  is not an admissible solution.

- ii) If  $p_* \leq 1/2$  and if, for example,  $E(X) = 0$  and  $m_0 = m$  then  $\vartheta_{**} = (2p^*, \alpha_*/2, \beta_*/2)$ . In such a case it would be crucial to build a conveniently constrained parametric space (most of the time a plot of the dataset helps in building reasonable constraints on the intercept and slope parameter spaces) expecting that it contains  $\vartheta_*$  but not  $\vartheta_{**}$ .
- iii) More generally one can expect that when the shape of the sample data (see *e.g.* Fig. 1) suggest that  $(m_0 - m)/2\beta_*\sigma_h^2$  is negligible with respect to  $\alpha_*$  and  $\beta_*$ , which occurs when  $m_0$  is close to  $m$  or/and  $\sigma_h^2\beta_*$  is very large, then the solution proposed in ii) is loosely speaking still valid.

## 5.2 Explicit formula of $H(\cdot, \vartheta)$ and its derivatives

In this section all the expressions are valid for all  $(\vartheta, y) \in \Theta \times \mathbb{R}$ , and the computation of the various derivative functions (under the integral sign) are all allowed according to Lebesgue's Theorem and condition (R). According to (8) and (10), we have

$$\begin{aligned} H(y, \vartheta) = & \frac{p^*}{p} \left( \int_{-\infty}^y \int_{\mathbb{R}} f(z + (\theta - \theta^*) \odot x) h(x) dx dz \right. \\ & + \left. \int_{-\infty}^{-y} \int_{\mathbb{R}} f(z + (\theta - \theta^*) \odot x) h(x) dx dz \right) \\ & + \frac{p - p^*}{p} \left( \int_{-\infty}^y \int_{\mathbb{R}} f_0(z + \theta \odot x) h(x) dx dz \right. \\ & + \left. \int_{-\infty}^{-y} \int_{\mathbb{R}} f_0(z + \theta \odot x) h(x) dx dz \right) - 1. \end{aligned}$$

For simplicity we introduce

$$\begin{aligned} F^\theta(y) &= \int_{-\infty}^y \int_{\mathbb{R}} f(z + (\theta - \theta^*) \odot x) h(x) dx dz, \\ F_0^\theta(y) &= \int_{-\infty}^y \int_{\mathbb{R}} f_0(z + \theta \odot x) h(x) dx dz, \end{aligned}$$

which leads to

$$\frac{\partial}{\partial p} H(y, \vartheta) = -\frac{p^*}{p^2} \left[ (F^\theta(y) + F^\theta(-y)) - (F_0^\theta(y) + F_0^\theta(-y)) \right].$$

Let us denote

$$\begin{aligned}\dot{F}^\alpha(y) &:= \frac{\partial}{\partial \alpha} F^\theta(y) = \int_{-\infty}^y \int_{\mathbb{R}} \dot{f}(z + (\theta - \theta_*) \odot x) h(x) dx dz, \\ \dot{F}_0^\alpha(y) &:= \frac{\partial}{\partial \alpha} F_0^\theta(y) = \int_{-\infty}^y \int_{\mathbb{R}} \dot{f}_0(z + \theta \odot x) h(x) dx dz,\end{aligned}$$

and for

$$\begin{aligned}\dot{F}^\beta(y) &:= \frac{\partial}{\partial \beta} F^\theta(y) = \int_{-\infty}^y \int_{\mathbb{R}} x \dot{f}(z + (\theta - \theta^*) \odot x) h(x) dx dz, \\ \dot{F}_0^\beta(y) &:= \frac{\partial}{\partial \beta} F_0^\theta(y) = \int_{-\infty}^y \int_{\mathbb{R}} x \dot{f}_0(z + \theta \odot x) h(x) dx dz,\end{aligned}$$

we obtain

$$\frac{\partial}{\partial \alpha} H(y, \vartheta) = \frac{p^*}{p} \left( \dot{F}^\alpha(y) + \dot{F}^\alpha(-y) \right) + \frac{p - p^*}{p} \left( \dot{F}_0^\alpha(y) + \dot{F}_0^\alpha(-y) \right).$$

$$\frac{\partial}{\partial \beta} H(y, \vartheta) = \frac{p^*}{p} \left( \dot{F}^\beta(y) + \dot{F}^\beta(-y) \right) + \frac{p - p^*}{p} \left( \dot{F}_0^\beta(y) + \dot{F}_0^\beta(-y) \right).$$

At point  $\vartheta = \vartheta_*$  the Hessian matrix of  $H(\cdot, \vartheta)$  defined in (34) is obtained by considering

$$\dot{H}(y, \vartheta_*) = \begin{pmatrix} \frac{1}{p} (F_0^{\theta_*}(y) + F_0^{\theta_*}(-y) - 1) \\ 2f(y) \\ 2f(y)E(X) \end{pmatrix}. \quad (48)$$

Let us denote now

$$\begin{aligned}\ddot{F}^{\beta, \alpha}(y) &= \frac{\partial^2}{\partial \beta \partial \alpha} F^\theta(y) = \int_{-\infty}^y \int_{\mathbb{R}} x \ddot{f}(z + (\theta - \theta_*) \odot x) h(x) dx dz, \\ \ddot{F}_0^{\beta, \alpha}(y) &= \frac{\partial^2}{\partial \beta \partial \alpha} F_0^\theta(y) = \int_{-\infty}^y \int_{\mathbb{R}} x \ddot{f}_0(z + \theta \odot x) h(x) dx dz, \\ \ddot{F}^{\alpha, \alpha}(y) &= \frac{\partial^2}{\partial \alpha^2} F^\theta(y) = \int_{-\infty}^y \int_{\mathbb{R}} \ddot{f}(z + (\theta - \theta_*) \odot x) h(x) dx dz, \\ \ddot{F}_0^{\beta, \beta}(y) &= \frac{\partial^2}{\partial \beta^2} F_0^\theta(y) = \int_{-\infty}^y \int_{\mathbb{R}} x^2 \ddot{f}_0(z + \theta \odot x) h(x) dx dz.\end{aligned}$$

We then obtain

$$\begin{aligned}
\frac{\partial^2}{\partial p^2} H(y, \vartheta) &= \frac{p^*}{p^3} [(F^\theta(y) + F^\theta(-y)) - (F_0^\theta(y) + F_0^\theta(-y))], \\
\frac{\partial^2}{\partial u \partial p} H(y, \vartheta) &= -\frac{p^*}{p^2} [(\dot{F}^u(y) + \dot{F}^u(-y)) - (\dot{F}_0^u(y) + \dot{F}_0^u(-y))], \quad u = \alpha, \beta, \\
\frac{\partial^2}{\partial u \partial v} H(y, \vartheta) &= \frac{p^*}{p} (\ddot{F}^{u,v}(y) + \ddot{F}^{u,v}(-y)) \\
&\quad + \frac{p - p^*}{p} (\ddot{F}_0^{u,v}(y) + \ddot{F}_0^{u,v}(-y)), \quad u = \alpha, \beta, \quad v = \alpha, \beta.
\end{aligned}$$

### 5.3 Boundedness

*Boundedness of  $\Psi_\theta(\cdot)$  and  $\Psi'_\theta(\cdot)$ .* If  $f$  and  $f_0$  are supposed to be bounded over  $\mathbb{R}$  then we clearly have from (8) that

$$|\Psi_\theta(y)| \leq \|f\|_\infty + \|f_0\|_\infty, \quad (\theta, y) \in \Phi \times \mathbb{R}.$$

The same kind of argument holds to prove boundedness of  $\dot{f}_\theta(\cdot)$  when (R) ii) is supposed.

*Boundedness of  $H(\cdot, \vartheta)$ .* Since for all  $\theta \in \Phi$  the functions  $F_\theta(\cdot)$  and  $J_\theta(\cdot)$  are both cdfs, we thus have, since  $\delta \leq p \leq 1 - \delta$ , from expression (12):

$$H(y, \vartheta) \leq \frac{4}{\delta} + 1, \quad (y, \vartheta) \in \mathbb{R} \times \Phi. \quad (49)$$

### 5.4 Integrable Lipschitz property of $\Psi_\theta(\cdot)$

From (8), for all  $(y, \theta, \theta') \in \mathbb{R} \times \Phi^2$  we have

$$\begin{aligned}
|\Psi_\theta(y) - \Psi_{\theta'}(y)| &\leq p_* \int_{\mathbb{R}} |f(y + (\theta - \theta_*) \odot x) - f(y + (\theta' - \theta_*) \odot x)| h(x) dx \\
&\quad + (1 - p_*) \int_{\mathbb{R}} |f_0(y + \theta \odot x) - f_0(y + \theta' \odot x)| h(x) dx. \quad (50)
\end{aligned}$$

Consider for simplicity the first integral term on the right hand side of (50) (the same argument holding for the second term). According to the Mean Value Theorem there exists, for all  $(x, y) \in \mathbb{R}^2$  and  $(\theta, \theta') \in \Phi^2$ , a value  $\gamma := \gamma(x, y, \theta, \theta')$  belonging to the line segment with extremities  $y + (\theta - \theta_*) \odot x$



and  $y + (\theta' - \theta_*) \odot x$ , or equivalently a  $\bar{\theta} := \bar{\theta}(x, y, \theta, \theta')$  belonging to the line segment with extremities  $\theta$  and  $\theta'$  such that  $\gamma = y + (\bar{\theta} - \theta_*) \odot x$  and

$$\begin{aligned}
& |f(y + (\theta - \theta_*) \odot x) - f(y + (\theta' - \theta_*) \odot x)| \\
&= |\dot{f}(\gamma)(\alpha - \alpha' + (\beta - \beta')x)| \\
&= |\dot{f}(y + (\bar{\theta} - \theta_*) \odot x)(\alpha - \alpha' + (\beta - \beta')x)| \\
&\leq \sup_{\theta \in \Phi} |\dot{f}(y + (\theta - \theta_*) \odot x)|(|\alpha - \alpha'| + |\beta - \beta'| |x|).
\end{aligned}$$

From condition (R) ii) there thus exists a nonnegative constant  $c$  such that

$$\begin{aligned}
& \int_{\mathbb{R}} |\Psi_{\theta}(y) - \Psi_{\theta'}(y)| \\
& \leq \sum_{j=0,1} (|\alpha - \alpha'|^{1-j} + |\beta - \beta'|^j) \\
& \int_{\mathbb{R} \times \mathbb{R}} |x|^j (\sup_{\theta \in \Phi} |\dot{f}(z + (\theta - \theta_*) \odot x)| + \sup_{\theta \in \Phi} |\dot{f}_0(z + \theta \odot x)|) h(x) dx dz \\
& \leq c \|\theta - \theta'\|_2.
\end{aligned}$$

## 5.5 Uniform Lipschitz property of $H(\cdot, \vartheta)$

Let us write

$$\begin{aligned}
& H(y, \vartheta) - H(y, \vartheta') \\
&= \frac{1}{p} (F_{\theta}(y) - F_{\theta'}(y) + F_{\theta}(-y) - F_{\theta'}(-y)) + \frac{p - p'}{pp'} (F_{\theta'}(y) + F_{\theta'}(-y)) \\
& \quad + \frac{1 - p}{p} (J_{\theta}(y) - J_{\theta'}(y) + J_{\theta}(-y) - J_{\theta'}(-y)) + \frac{p - p'}{pp'} (J_{\theta'}(y) + J_{\theta'}(-y)).
\end{aligned}$$

To prove the uniform Lipschitz property of  $H(\cdot, \vartheta)$  we need to prove it for  $J_{\theta}(\cdot)$  and  $F_{\theta}(\cdot)$ . We begin with the simplest term  $J_{\theta}(\cdot)$ . According again to the mean value theorem, for all  $y \in \mathbb{R}$ , all  $(x, z) \in \mathbb{R}^2$  with  $z \leq y$ , and all  $(\theta, \theta') \in \Phi^2$  there exists  $\bar{\theta} := \bar{\theta}(x, z, \theta, \theta')$  belonging to the line segment with

extremities  $\theta$  and  $\theta'$  such that

$$\begin{aligned}
|J_\theta(y) - J_{\theta'}(y)| &\leq \int_{-\infty}^y \int_{\mathbb{R}} |f_0(z + \theta \odot x) - f_0(z + (\theta' \odot x))| h(x) dx dz \\
&= \int_{-\infty}^y \int_{\mathbb{R}} |\dot{f}_0(z + \bar{\theta} \odot x)(|\alpha - \alpha'| + |\theta - \theta'| |x|) h(x) dx dz, \\
&\leq \int_{-\infty}^y \int_{\mathbb{R}} \sup_{\theta \in \Phi} |\dot{f}_0(z + \theta \odot x) h(x) dx dz| |\alpha - \alpha'| \\
&\quad + \int_{-\infty}^y \int_{\mathbb{R}} |x| \sup_{\theta \in \Phi} |\dot{f}_0(z + ux) h(x) dx dz| |\beta - \beta'| \\
&\leq c \|\theta - \theta'\|_2,
\end{aligned}$$

where  $c$  denotes a nonnegative constant arising from condition (R) ii). Using the same kind of argument we prove that there exists a nonnegative constant  $c'$  such that for all  $(y, \theta, \theta') \in \mathbb{R} \times \Phi^2$

$$|F_\theta(y) - F_{\theta'}(y)| \leq c' \|\theta - \theta'\|_2.$$

In conclusion, for all  $y \in \mathbb{R}$ , there exists a nonnegative constant  $c''$  such that for all  $(y, \vartheta, \vartheta') \in \mathbb{R} \times \Theta^2$

$$|H(y, \vartheta) - H(y, \vartheta')| \leq \frac{2}{\delta} (c + c') \|\theta - \theta'\|_2 + 4 \frac{|p - p'|}{\delta^2} \leq c'' \|\vartheta - \vartheta'\|_3.$$

## 5.6 Uniform almost sure rate of convergence of $d_n$

Let us consider

$$|d_n(\vartheta) - d(\vartheta)| \leq T_{1,n}(\vartheta) + T_{2,n}(\vartheta),$$

where

$$\begin{aligned}
T_{1,n}(\vartheta) &:= \left| \frac{1}{n} \sum_{i=1}^n H^2(V_i; \vartheta, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta}) - H^2(V_i; \vartheta, F_\theta, J_\theta) \right|, \\
T_{2,n}(\vartheta) &:= \left| \frac{1}{n} \sum_{i=1}^n H^2(V_i; \vartheta, F_\theta, J_\theta) - E(H^2(V_1; \vartheta, F_\theta, J_\theta)) \right|.
\end{aligned}$$

*Uniform almost sure rate of convergence of  $T_{1,n}$ .* Note first that from boundedness of  $H(\cdot; \vartheta, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta})$  and  $H(\cdot; \vartheta, F_\theta, J_\theta)$  given by (49), there exist non-

negative constants  $C$  and  $C'$  such that

$$\begin{aligned}
T_{1,n}(\vartheta) &= \left| \frac{1}{n} \sum_{i=1}^n (H(V_i; \vartheta, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta}) + H(V_i; \vartheta, F_\theta, J_\theta)) \right. \\
&\quad \left. \times (H(V_i; \vartheta, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta}) - H(V_i; \vartheta, F_\theta, J_\theta)) \right| \\
&\leq C \sup_{y \in \mathbb{R}} |H(y; \vartheta, \tilde{F}_{n,\theta}, \hat{J}_{n,\theta}) - H(y; \vartheta, F_\theta, J_\theta)| \\
&\leq C' \left( \sup_{y \in \mathbb{R}} |\hat{J}_{n,\theta}(y) - J_\theta(y)| + \sup_{y \in \mathbb{R}} |\tilde{F}_{n,\theta}(y) - F_\theta(y)| \right).
\end{aligned}$$

Let us now denote

$$\begin{aligned}
T_{1,n}^{(1)} &:= \sup_{\theta \in \Phi} \sup_{y \in \mathbb{R}} |\hat{J}_{n,\theta}(y) - J_\theta(y)|, \\
T_{1,n}^{(2)} &:= \sup_{\theta \in \Phi} \sup_{y \in \mathbb{R}} |\tilde{F}_{n,\theta}(y) - F_\theta(y)|.
\end{aligned}$$

*Convergence rate of  $T_{1,n}^{(1)}$ .* For simplicity we will suppose that  $\text{proj}_2(\Phi) \subset [0, A]$ , where  $A$  is a nonnegative real number and for all  $(x, y) \in \mathbb{R}^2$ ,  $\text{proj}_2 : (x, y) \mapsto y$ . Let us introduce  $P_n^X = n^{-1} \sum_{i=1}^n \delta_{X_i}$  the empirical measure associated to the iid sample  $(X_1, \dots, X_n)$  with common probability distribution  $P^X$  with pdf and cdf resp. denoted by  $h$  and  $H$ ). We use the fonctionnal notation  $Pf = \int f dP$ . Notice now that, according to expression (17), we have for all  $y \in \mathbb{R}$ ,

$$\begin{aligned}
\hat{J}_{n,\theta}(y) - J_\theta(y) &= \frac{1}{n} \sum_{i=1}^n F_0(y + \alpha + \beta X_i) - E(F_0(y + \alpha + \beta X)) \quad (51) \\
&= (P_n^X - P^X)F_0(y + \alpha + \beta \cdot).
\end{aligned}$$

Let consider the class of functions

$$\mathcal{F}_0 = \{x \mapsto F_0(u + \beta x); \quad u \in \mathbb{R}, \quad \beta \in [0, A]\}.$$

Since

$$(P_n^X - P^X)F_0(u + \beta \cdot) = (P_n^X - P^X)F_0(y + \beta(\cdot \vee 0)) + (P_n^X - P^X)F_0(y + \beta(\cdot \wedge 0)),$$

it is enough to study the empirical process indexed by the classes of functions

$$\begin{aligned}
\mathcal{F}_0^+ &= \{x \mapsto F_0(u + \beta(x \vee 0)); \quad u \in \mathbb{R}, \quad \beta \in [0, A]\}, \\
\mathcal{F}_0^- &= \{x \mapsto F_0(u + \beta(x \wedge 0)); \quad u \in \mathbb{R}, \quad \beta \in [0, A]\}.
\end{aligned}$$

For simplicity we denote  $\Gamma_{y,\alpha}(x) = F_0(y + \alpha(x \vee 0))$  and only consider the class  $\mathcal{F}_0^+$ , the class  $\mathcal{F}_0^-$  being treated in a entirely same way. Since  $F_0$  is a cdf, for  $\beta_1 \leq \beta \leq \beta_2$  and  $u_1 \leq u \leq u_2$  we have

$$\Gamma_{u_1,\beta_1}(x) \leq \Gamma_{u,\beta}(x) \leq \Gamma_{u_2,\beta_2}(x), \quad x \in \mathbb{R},$$

and, since  $F_0$  is supposed to be Lipschitz,

$$0 \leq \Gamma_{u_2,\beta_2}(x) - \Gamma_{u_1,\beta_1}(x) \leq C(u_2 - u_1 + (\beta_2 - \beta_1)(x \vee 0)).$$

Let consider now  $\varepsilon > 0$ , and  $(\overline{u}_\varepsilon, \underline{u}_\varepsilon) \in \mathbb{R}^2$  such that

$$F_0(\overline{u}_\varepsilon) \geq 1 - \varepsilon, \quad \text{and} \quad F_0(\underline{u}_\varepsilon) \leq \varepsilon.$$

Note that  $\overline{u}_\varepsilon$  and  $\underline{u}_\varepsilon$  do not depend on  $\beta$ . For all  $N \in \mathbb{N}$ , define

$$\underline{u}_\varepsilon = u_{1,\varepsilon} \leq u_{2,\varepsilon} \leq \dots \leq u_{N,\varepsilon} = \overline{u}_\varepsilon,$$

and consider  $N(\varepsilon)$  the smallest integer such that  $u_{i,\varepsilon} - u_{i-1,\varepsilon} \leq \varepsilon$  for  $i = 2, \dots, N(\varepsilon)$ . We denote by  $\lceil \cdot \rceil$  the integer part function. For all  $\varepsilon$  small enough we clearly have

$$N(\varepsilon) \leq \left\lceil \frac{\overline{u}_\varepsilon - \underline{u}_\varepsilon}{\varepsilon} \right\rceil \leq 2 \frac{\overline{u}_\varepsilon - \underline{u}_\varepsilon}{\varepsilon}.$$

Let us now define  $\alpha_{i,\varepsilon} = \varepsilon(i - 1)$ ,  $i = 1, \dots, M(\varepsilon)$ , where  $M(\varepsilon) = \lceil [A + 1]/\varepsilon \rceil$  and thus  $\alpha_{M(\varepsilon),\varepsilon} > A$ . Observe in addition that

$$\begin{aligned} \|\Gamma_{u_{i+1,\varepsilon},\beta_{j+1,\varepsilon}} - \Gamma_{u_{i,\varepsilon},\beta_{j,\varepsilon}}\|_{2,P^X}^2 &= c^2 E((u_{i+1,\varepsilon} - u_{i,\varepsilon} + (\beta_{j+1,\varepsilon} - \beta_{j,\varepsilon})(X_1 \wedge 0))^2) \\ &\leq 2c^2 \varepsilon^2 + 2C^2 \varepsilon E(X_1^2) \\ &= 2c^2 \varepsilon^2 (1 + E(X_1^2)). \end{aligned}$$

Hence the expression

$$[\Gamma_{u_{i+1,\varepsilon},\beta_{j+1,\varepsilon}} - \Gamma_{u_{i,\varepsilon},\beta_{j,\varepsilon}}], \quad 1 \leq i \leq N(\varepsilon), \quad 1 \leq j \leq M(\varepsilon),$$

is a  $\left(c\sqrt{2(1 + E(X_1^2))}\right)$ -covering of  $\mathcal{F}_0^+$  in the  $L_2(P^X)$ -norm sense. Using the standard notation  $N_{[]}(\cdot)$  (see van der Vaart and Wellner [25]) the covering number of the class  $\mathcal{F}_0^+$  is bounded as follows

$$N_{[]}(\varepsilon, \mathcal{F}_0^+, L_2(P^X)) \leq cN(\varepsilon)M(\varepsilon) \leq c' \frac{\overline{u}_\varepsilon - \underline{u}_\varepsilon}{\varepsilon^2}.$$

Thus if there exist constants  $C$  and  $V$  such that

$$|\overline{u_\varepsilon}| \wedge |\underline{u_\varepsilon}| \leq C/\varepsilon^V, \quad (52)$$

we get  $N(\varepsilon)M(\varepsilon) \leq C/\varepsilon^{V+2}$  which allows us to use Theorem 2.14.9, p. 246 in [25] since their Condition (2.14.7), p. 245 is then satisfied after replacing their constant  $V$  by  $V + 2$ . Let us discuss condition (52). For  $\varepsilon$  small enough this condition is true if  $\overline{y_\varepsilon} \leq C/\varepsilon^V$  and  $\underline{u_\varepsilon} \geq -C/\varepsilon^V$ . Denoting by  $F_0^{\leftarrow}$  the quantile function of  $F_0$ , condition (52) becomes

$$F_0^{\leftarrow}(1 - \varepsilon) \leq C/\varepsilon^V \quad \text{and} \quad F_0^{\leftarrow}(\varepsilon) \geq -C/\varepsilon^V. \quad (53)$$

We consider for simplicity the first condition in (53) (the second one being treated in the same way); it is equivalent to  $F_0(C/\varepsilon^V) \geq 1 - \varepsilon$ , and taking  $t = C/\varepsilon^V$  this condition turns into

$$F_0(t) \geq 1 - C/t^{1/V}. \quad (54)$$

Thus it suffices to have

$$\liminf_{t \rightarrow \infty} \frac{-\log(1 - F_0(t))}{\log(t)} > 0. \quad (55)$$

Finally, using the symmetry of  $f_0$ , condition (52) holds if

$$\liminf_{t \rightarrow \infty} \frac{-2 \log F_0(-t)}{\log(t)} > 0, \quad (56)$$

which is insured by condition (R) vi). In conclusion if (56) is satisfied and  $E(X_1^2) < \infty$  then, according to Theorem 2.14.16, p. 248 in van der Vaart and Wellner [25], we obtain

$$\sup_{\theta \in \Phi} \|\hat{J}_{n,\theta} - J_\theta\|_\infty \leq \|P_n^X - P^X\|_{\mathcal{F}_0} = o_{a.s.}(n^{-1/2+\gamma}), \quad \gamma > 0.$$

*Convergence rate of  $T_{1,n}^{(2)}$ .* Recall that  $F_\theta$  is the cdf of  $Y_i - \theta \odot X_i$ , i.e.

$$F_\theta(y) = P(Y_i - \theta \odot X_i \leq y),$$

and

$$F_{n,\theta}(y) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{Y_i - \theta \odot X_i \leq y}.$$

Let  $K$  a kernel satyving (K). The  $K$ -regularized versions of  $F_\theta$  and  $F_{n,\theta}$  are

$$\tilde{F}_\theta = K * F_\theta, \quad \tilde{F}_{n,\theta} = K * F_{n,\theta}.$$

Let us denote by  $P_n^{X,Y}$  the empirical measure

$$P_n^{X,Y} = \frac{1}{n} \sum_{i=1}^n \delta_{X_i, Y_i},$$

and by  $P^{X,Y}$  the law of  $(X_1, Y_1)$ .

The set of functions for which  $(x, y) \mapsto ax + by + c$  being a 3-dimensionnal vector space, Corollary 2.5 in Kuelbs and Dudley [19] shows that the class of sets

$$\mathcal{C} = \left\{ \left\{ (u, v) \in \mathbb{R}^2 : au + bv + c < 0 \right\} ; (a, b, c) \in \mathbb{R}^3 \right\},$$

is a Strassen log-log class, which implies that *a.s.*

$$\limsup_{n \rightarrow \infty} \sup_{\mathcal{C}} \sqrt{\frac{n}{2 \log \log(n)}} (P_n^Z - P^Z)(C) = \sup_{C \in \mathcal{C}} \sqrt{P^Z(C)(1 - P^Z(C))} \leq 1/2.$$

Since  $\mathcal{C}$  contains the class

$$\mathcal{S} := \left\{ \left\{ (u, v) \in \mathbb{R}^2 : v - (\alpha + \beta u) < y \right\} ; (\alpha, \beta; y) \in \Phi \times \mathbb{R} \right\},$$

it follows that, for all set  $S \in \mathcal{S}$ ,  $P^Z(S) = \int_S dP^{X,Y}(u, v) = P(Y - (\alpha + \beta X) < y) = F_\theta(y)$  and for the same reason  $P_n^Z(S) = F_{n,\theta}(y)$ , we have

$$\limsup_{n \rightarrow \infty} \sup_{(\theta, y) \in \Phi \times \mathbb{R}} \sqrt{\frac{n}{\log \log(n)}} (F_{n,\theta} - F_\theta)(y) \leq 1/2 \quad a.s. \quad (57)$$

Now if we replace  $F_{n,\theta}$  by its regularized version  $\tilde{F}_{n,\theta}$  the approximation is controlled as follows,

$$\begin{aligned} \tilde{F}_{n,\theta}(y) - F_{n,\theta}(y) &= \tilde{F}_{n,\theta}(y) - E(\tilde{F}_{n,\theta}(y)) + E(\tilde{F}_{n,\theta}(y)) - F_\theta(y) + F_\theta(y) - F_{n,\theta}(y) \\ &= \tilde{F}_{n,\theta}(y) - E(\tilde{F}_{n,\theta}(y)) - [F_{n,\theta}(y) - E(F_{n,\theta}(y))] \\ &\quad + E(\tilde{F}_{n,\theta}(y)) - F_\theta(y), \end{aligned} \quad (58)$$

recalling that  $E(F_{n,\theta}(y)) = F_\theta(y)$ . The first term on the right hand side of (58) satisfies

$$\begin{aligned}\tilde{F}_{n,\theta}(y) - E(\tilde{F}_{n,\theta}(y)) &= \int_{\mathbb{R}} K(y-u) d(F_{n,\theta} - E(F_{n,\theta}))(u) \\ &= \int_{\mathbb{R}} (F_{n,\theta} - E(F_{n,\theta}))(u) dK(y-u) \\ &= \int_{\mathbb{R}} (F_{n,\theta} - E(F_{n,\theta}))(y-s) dK(s).\end{aligned}$$

Thus, if we denote  $\Delta_{n,\theta}(y) := F_{n,\theta}(y) - E(F_{n,\theta}(y)) = F_{n,\theta}(y) - F_\theta(y)$ , we obtain

$$\begin{aligned}\left| \tilde{F}_{n,\theta}(y) - E(\tilde{F}_{n,\theta}(y)) - [F_{n,\theta}(y) - E(F_{n,\theta}(y))] \right| \\ \leq \left| \int_{\mathbb{R}} (\Delta_{n,\theta}(y-s) - \Delta_{n,\theta}(y)) dK(s) \right| \\ \leq \sup_{(\theta,y) \in \Phi \times \mathbb{R}} |\Delta_{n,\theta}(y)| \|K\|_{TV}.\end{aligned}$$

The last bias-term on the right hand side of (58) can be studied using the  $R_{2n}$  bound in [22], p. 766, equation (e), which establishes that for each  $\theta \in \Phi$

$$\sup_{y \in \mathbb{R}} |E(\tilde{F}_{n,\theta}) - F_\theta|(y) \leq \frac{\|\dot{f}_\theta\|_\infty k_2}{2}. \quad (59)$$

If  $K$  is replaced by  $K_n(\cdot) = K(\cdot/b_n)$  and we let  $k_{2,n} := \int_{\mathbb{R}} x^2 dK_n(x) = b_n^2 k_2$ , then (57–59) lead to

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{n}{\log \log(n)}} \sup_{(\theta,y) \in \Phi \times \mathbb{R}} (F_{n,\theta} - F_\theta)(y) < \infty \quad a.s., \quad (60)$$

whenever  $\limsup(n/\log \log(n))^{1/2} k_{2,n} < \infty$  which holds when

$$\limsup_{n \rightarrow \infty} \sqrt{\frac{n}{\log \log(n)}} b_n^2 < \infty, \quad (61)$$

and  $\sup_{\theta \in \Phi} \|\dot{f}_\theta\|_\infty < \infty$  which has been proved in Section 5.3 under Condition (R) ii).

*Uniform almost sure rate of convergence of  $T_{2,n}$ .* Considering for all  $i \geq 0$ , the random variable  $W_i(\vartheta) := H^2(V_i; \vartheta)$ , where  $\vartheta \in \Theta$ , we see that

$$\sup_{\vartheta \in \Theta} T_{2,n}(\vartheta) = \sup_{\vartheta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^n W_i(\vartheta) - E(W_i(\vartheta)) \right|,$$

where the right hand term is the supremum of an empirical process indexed by a class of Lipschitz bounded functions, which is known to be  $o_{a.s.}(n^{-1/2+\gamma})$  for all  $\gamma > 0$  (see [2], for details), which concludes the proof.

**Acknowledgments.** The author thanks the referees for their helpful and constructive comments. He is also very grateful to Philippe Barbe and David Hunter for their help and good advice during the writing of this manuscript.

## References

- [1] BENJAMINI, Y., and HOCHBERG, Y. (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Royal Statist. Soc. Ser. B*, **57**, 289–300.
- [2] BORDES, L., MOTTELET, S. and VANDEKERKHOVE, P. (2006). Semiparametric estimation of a two-component mixture model. *Ann. Statist.*, **34**, 1204–1232.
- [3] BORDES, L., DELMAS, C. and VANDEKERKHOVE, P. (2006). Semiparametric estimation of a two-component mixture model when a component is known. *Scand. J. Statist.*, **33**, 733–752.
- [4] BORDES, L., CHAUVEAU, D. and VANDEKERKHOVE, P. (2007). A stochastic EM algorithm for a semiparametric mixture model. *Comput. Statist. Data Anal.*, **51**, 5429–5443.
- [5] BORDES, L., and VANDEKERKHOVE, P. (2010). Semiparametric Two-Component Mixture Model with a Known Component: an Asymptotically Normal Estimator. *Math. Method. Statist.*, **19**, 22–41.
- [6] BOUYEYRON C. AND JACQUES J. (2010). Adaptive mixtures of regressions: Improving predictive inference when population has changed. *Pattern Recognition Letters*, **31**, 2237–2247.



- [7] BOWMAN, A.W. and AZZALINI, A. (2003). Computational aspects of nonparametric smoothing with illustrations from the sm library. *Comp. Statist. Data Anal.*, **42**, 545–56.
- [8] CRUZ-MEDINA, I. R. and HETTMANSPERGER, T. P. (2004). Non-parametric estimation in semiparametric univariate mixture models. *J. Statist. Comput. Simulation*, **74**, 513–524.
- [9] Cuzick, J. (1992) Semiparametric additive regression. *J. Royal Statist. Soc. Ser. B*, **3**, 831–843.
- [10] Cuzick, J. (1992) Efficient estimates in semiparametric additive regression models with unknown error distribution. *Ann. Statist.*, **20**, 1129–1136.
- [11] Devroye, L. (1983) The equivalence of weak, strong and complete convergence in  $L_1$  for kernel density estimates. *Ann. Statist.*, **11**, 896–904.
- [12] EFRON, B. (2007). Size, power and false discovery rate. *Ann. Statist.*, **35**, 1351–1377.
- [13] HALL, P. and ZHOU, X. H. (2003). Nonparametric estimation of component distributions in a multivariate mixture. *Ann. Statist.*, **31**, 201–224.
- [14] HENG-YAN LEUNG, D. and QIN, J. (2006). Semi-parametric inference in bivariate (multivariate) mixture model. *Statist. Sinica.*, **16**, 153–163.
- [15] HETTMANSPERGER, T. P. and THOMAS, H. (2000). Almost nonparametric inference for repeated measures in mixture models. *J. Royal Statist. Soc. Ser. B*, **62** 811–825.
- [16] HOHMANN, D. and HOLZMANN, H. (2011). Semiparametric location mixtures with distinct components. *Preprint*.
- [17] HUNTER, D. R., WANG, S. and HETTMANSPERGER, T. P. (2007). Inference for mixtures of symmetric distributions. *Ann. Statist.*, **35**, 224–251.

- [18] KUI W. and YAU, K. K. W. (2007) Two-component Poisson mixture regression modelling of count data with bivariate random effects. *Math. Comput. Modelling*, **46**, 1468–1476.
- [19] KUELBS, J. and DUDLEY, R. M. (1980). Log Log Laws for Empirical Measures *Ann. Probab.*, **8**, 405–418.
- [20] MAIBORODA, R. and SUGAKOVA, O. (2010). Adaptive estimating equations for a location parameter constructed by using observations with admixture. *Th. Probab. Math. Statist.*, **80**, 101–110.
- [21] MARTIN-MAGNIETTE, M-L., MARY-HUARD, T., BÉRARD C. and ROBIN, S. (2008) ChIPmix: mixture model of regressions for two-color ChIP-chip analysis. *Bioinformatics*, **24**, 181–186.
- [22] SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- [23] SCHUSTER, E. F. and BARKER, R. C. (1987). Using the bootstrap in testing symmetry versus asymmetry. *Comm. Statist. Simulation Comput.*, **16**, 69–84.
- [24] TITTERINGTON, D. M., SMITH, A. F. M. and MAKOV, U. E. (1985). *Statistical Analysis of Finite Mixture Distributions*, Wiley, Chichester.
- [25] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer-Verlag, New-York.
- [26] YAU, K. K. W. , LEEB, A. H., and NG A. S. K. (2003) Finite mixture regression model with random effects: application to neonatal hospital length of stay. *Comput. Statist. Data Anal.*, **41**, 359–366.
- [27] YOUNG, D. S, and HUNTER D. R. (2010). Mixtures of Regressions with Predictor-Dependent Mixing Proportions. *Comput. Statist. Data Anal.*, **54**, 2253–2266.

- [28] YU K., MAMMEN E. and PARK B. U. (2010) Semiparametric Regression: Efficiency Gains From Modeling the Nonparametric Part. *To appear in Bernoulli*.
- [29] ZHU H. T. and ZHANG H. (2004) Hypothesis testing in mixture regression models. *J. Royal Statist. Soc. Ser. B*, **66**, 3–16.